

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US2005/010912

International filing date: 31 March 2005 (31.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/559,042
Filing date: 01 April 2004 (01.04.2004)

Date of receipt at the International Bureau: 12 August 2005 (12.08.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

10029



THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 02, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: 60/559,042

FILING DATE: April 01, 2004

RELATED PCT APPLICATION NUMBER: PCT/US05/10912



Certified by

Don W. Dudas

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

040104
01919 U.S. PTO

Approved for use through 7/31/2008. OMB 0651-0032
U.S. Patent and Trademark Office, U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. EL 961008104 US

22857 - U.S. PTO
601559042

040104

INVENTOR(S)			
Given Name (first and middle (if any))	Family Name or Surname	Residence (City and either State or Foreign Country)	
Steven	MAH	San Diego, California	
Andreas	BRAUN	San Diego, California	
Stefan M.	KAMMERER	San Diego, California	
Additional inventors are being named on the <u>1</u> separately numbered sheets attached hereto			
TITLE OF THE INVENTION (500 characters max)			
METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND TREATMENTS THEREOF			
Direct all correspondence to: CORRESPONDENCE ADDRESS			
<input checked="" type="checkbox"/> Customer Number: <u>25225</u> OR <input type="checkbox"/> Firm or Individual Name			
Address			
City	State	Zip	
Country	Telephone	Fax	
ENCLOSED APPLICATION PARTS (check all that apply)			
<input checked="" type="checkbox"/> Specification Number of Pages <u>162</u>		<input type="checkbox"/> CD(s), Number <u> </u>	
<input checked="" type="checkbox"/> Drawing(s) Number of Sheets <u>1</u>		<input checked="" type="checkbox"/> Other <u> </u>	
<input checked="" type="checkbox"/> Application Data Sheet. See 37 CFR 1.76 (4 pages) (specify):		Return Receipt Postcard	
METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT			
<input checked="" type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27. <input type="checkbox"/> A check or money order is enclosed to cover the filing fees. <input checked="" type="checkbox"/> The Director is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: <u>03-1952</u> <input type="checkbox"/> Payment by credit card. Form PTO-2038 is attached.			
			FILING FEE AMOUNT (\$) <u>80.00</u>
The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.			
<input checked="" type="checkbox"/> No <input type="checkbox"/> Yes, the name of the U.S. Government agency and the Government contract number are:			

Respectfully submitted,

[Page 1 of 2]

Date April 1, 2004

SIGNATURE

TYPED OR
PRINTED NAME

Bruce D. Grant

TELEPHONE

(858) 720-7962

REGISTRATION NO.
(if appropriate)

47,608

Docket Number:

524593008900

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

I hereby certify that this correspondence is being deposited with the U.S. Postal Service as Express Mail, Airbill No. EL 961008104 US, in an envelope addressed to: Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date shown below.

Dated: 4/1/04

Signature: Deborah Wykes

(Deborah Wykes)

Additional Page

Approved for use through 07/31/06. OMB 0651-0032

U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Docket Number 524593008900

INVENTOR(S)/APPLICANT(S)

Given Name (first and middle if any)	Family or Surname	Residence (City and either State or Foreign Country)
Matthew Roberts	NELSON	San Marcos, California
Rikard Henry	RENELAND	San Diego, California
Maria L.	LANGDOWN	San Diego, California

METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND TREATMENTS THEREOF

Field of the Invention

[0001] The invention relates to genetic methods for identifying risk of osteoarthritis and treatments that specifically target such diseases.

Background

[0002] Osteoarthritis (OA) is a chronic disease usually affecting weight-bearing synovial joints. There are approximately 20 million Americans affected by OA and it is the leading cause of disability in the United States. In addition to extensive human suffering, OA also accounts for nearly all knee replacements and more than half of all hip replacements in the United States. Despite its prevalence, OA is poorly understood and there are few treatments available besides anti-inflammatory drugs and joint replacement.

[0003] Most commonly affecting middle-aged and older people, OA can range from very mild to very severe. It affects hands and weight-bearing joints such as knees, hips, feet and the back. Knee OA can be as disabling as any cardiovascular disease except stroke.

[0004] OA is characterized by the breakdown of cartilage in joints. Cartilage in joints cushions the ends of bones, and cartilage breakdown causes bones to rub against each other, causing pain and loss of movement. Type II collagen is the main component of cartilage, comprising 15-25% of the wet weight, approximately half the dry weight, and representing 90-95% of the total collagen content in the tissue. It forms fibrils that endow cartilage with tensile strength (Mayne, R. Arthritis Rheum. 32:241-246 (1989)).

Summary

[0005] It has been discovered that certain polymorphic variations in human genomic DNA are associated with osteoarthritis. In particular, polymorphic variants in a locus containing a *ADAMTS2* region in human genomic DNA have been associated with risk of osteoarthritis.

[0006] Thus, featured herein are methods for identifying a subject at risk of osteoarthritis and/or a risk of osteoarthritis in a subject, which comprise detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in or around the loci described herein in a human nucleic acid sample. In an embodiment, two or more polymorphic variations are detected in two or more regions of which one is the *ADAMTS2* region. In certain embodiments, 3 or more, or 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 or more polymorphic variants are detected.

[0007] Also featured are nucleic acids that include one or more polymorphic variations associated with occurrence of osteoarthritis, as well as polypeptides encoded by these nucleic acids. In addition, provided are methods for identifying candidate therapeutic molecules for treating osteoarthritis, as well as methods for treating osteoarthritis in a subject by identifying a subject at risk of osteoarthritis and treating the subject with a suitable prophylactic, treatment or therapeutic molecule.

[0008] Also provided are compositions comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and/or a *ADAMTS2* nucleic acid, with a RNAi, siRNA, antisense DNA or RNA, or ribozyme nucleic acid designed from a *ADAMTS2* nucleotide sequence. In an embodiment, the RNAi, siRNA, antisense DNA or RNA, or ribozyme nucleic acid is designed from a *ADAMTS2* nucleotide sequence that includes one or more polymorphic variations associated with osteoarthritis, and in some instances, specifically interacts with such a nucleotide sequence. Further, provided are arrays of nucleic acids bound to a solid surface, in which one or more nucleic acid molecules of the array have a *ADAMTS2* nucleotide sequence, or a fragment or substantially identical nucleic acid thereof, or a complementary nucleic acid of the foregoing. Featured also are compositions comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and/or a *ADAMTS2* polypeptide, with an antibody that specifically binds to the polypeptide. Thus, featured is an antibody that specifically binds to an epitope in the polypeptide that includes an amino acid encoded by a polymorphic site associated with osteoarthritis. In certain embodiments, the antibody specifically binds to an epitope comprising a valine or isoleucine encoded by rs398829 (e.g., an antibody that binds to an epitope comprising a valine at position 245 in an *ADAMTS2* polypeptide). A valine at position 245 is associated with increased risk of osteoarthritis.

Brief Description of the Drawings

[0009] Figure 1 shows proximal SNPs in a *ADAMTS2* region in genomic DNA. The position of each SNP in the chromosome is shown on the x-axis and the y-axis provides the negative logarithm of the p-value comparing the estimated allele to that of the control group. Also shown in the figure are exons and introns of the region in the approximate chromosomal positions.

Detailed Description

[0010] It has been discovered that a polymorphic variant in a locus containing a *ADAMTS2* region is associated with occurrence of osteoarthritis in subjects. Thus, detecting genetic determinants associated with an increased risk of osteoarthritis occurrence can lead to early identification of a predisposition to osteoarthritis and early prescription of preventative measures. Also, associating a *ADAMTS2* polymorphic variant with osteoarthritis has provided new targets for screening molecules useful in treatments of osteoarthritis.

Osteoarthritis and Sample Selection

[0011] Osteoarthritis (OA), or degenerative joint disease, is one of the oldest and most common types of arthritis. It is characterized by the breakdown of the joint's cartilage. Cartilage is the part of the joint that cushions the ends of bones, and its breakdown causes bones to rub against each other, causing pain and loss of movement. Type II collagen is the main component of cartilage, comprising 15-25% of the wet weight, approximately half the dry weight, and representing 90-95% of the total collagen content in the tissue. It forms fibrils that endow cartilage with tensile strength (Mayne, R. Arthritis Rheum. 32:241-246 (1989)).

[0012] Most commonly affecting middle-aged and older people, OA can range from very mild to very severe. It affects hands and weight-bearing joints such as knees, hips, feet and the back. Knee OA can be as disabling as any cardiovascular disease except stroke. Whereas Ehlers-Danlos syndrome type VIIC is characterized by the retention of the N-terminal propeptide of type I collagen, osteoarthritis has been characterized by increased levels of type II collagen in osteoarthritic cartilage as measured by elevated C-propeptide concentrations (Nelson et al. (1998) J. Clin. Invest. 102(12):2115-2125).

[0013] Osteoarthritis affects an estimated 20.7 million Americans, mostly after age 45, with women more commonly affected than men. Physicians make a diagnosis of OA based on a physical exam and history of symptoms. X-rays are used to confirm diagnosis. Most people over 60 reflect the disease on X-ray, and about one-third have actual symptoms.

[0014] There are many factors that can cause OA. Obesity may lead to osteoarthritis of the knees. In addition, people with joint injuries due to sports, work-related activity or accidents may be at increased risk of developing OA.

[0015] Genetics has a role in the development of OA. Some people may be born with defective cartilage or with slight defects in the way that joints fit together. As a person ages, these defects may cause early cartilage breakdown in the joint or the inability to repair damaged or deteriorated cartilage in the joint.

[0016] Inclusion or exclusion of samples for an osteoarthritis pool may be based upon the following criteria: ethnicity (e.g., samples derived from an individual characterized as Caucasian); parental ethnicity (e.g., samples derived from an individual of British paternal and maternal descent); relevant phenotype information for the individual (e.g., case samples derived from individuals diagnosed with specific knee osteoarthritis (OA) and were recruited from an OA knee replacement clinic). Control samples may be selected based on relevant phenotype information for the individual (e.g., derived from individuals free of OA at several sites (knee, hand, hip etc)); and no family history of OA and/or rheumatoid arthritis. Additional phenotype information collected for both cases and controls may include age of the individual, gender, family history of OA, diagnosis with osteoarthritis (joint location of OA, date of primary diagnosis, age of individual as of primary diagnosis), knee history (current symptoms,

any major knee injury, meniscectomy, knee replacement surgery, age of surgery), HRT history, osteoporosis diagnosis.

[0017] Based in part upon selection criteria set forth above, individuals having osteoarthritis can be selected for genetic studies. Also, individuals having no history of osteoarthritis often are selected for genetic studies, as described hereafter.

Polymorphic Variants Associated with Osteoarthritis

[0018] A genetic analysis provided herein linked osteoarthritis with polymorphic variant nucleic acid sequences in the human genome. As used herein, the term “polymorphic site” refers to a region in a nucleic acid at which two or more alternative nucleotide sequences are observed in a significant number of nucleic acid samples from a population of individuals. A polymorphic site may be a nucleotide sequence of two or more nucleotides, an inserted nucleotide or nucleotide sequence, a deleted nucleotide or nucleotide sequence, or a microsatellite, for example. A polymorphic site that is two or more nucleotides in length may be 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more, 20 or more, 30 or more, 50 or more, 75 or more, 100 or more, 500 or more, or about 1000 nucleotides in length, where all or some of the nucleotide sequences differ within the region. A polymorphic site is often one nucleotide in length, which is referred to herein as a “single nucleotide polymorphism” or a “SNP.”

[0019] Where there are two, three, or four alternative nucleotide sequences at a polymorphic site, each nucleotide sequence is referred to as a “polymorphic variant” or “nucleic acid variant.” Where two polymorphic variants exist, for example, the polymorphic variant represented in a minority of samples from a population is sometimes referred to as a “minor allele” and the polymorphic variant that is more prevalently represented is sometimes referred to as a “major allele.” Many organisms possess a copy of each chromosome (*e.g.*, humans), and those individuals who possess two major alleles or two minor alleles are often referred to as being “homozygous” with respect to the polymorphism, and those individuals who possess one major allele and one minor allele are normally referred to as being “heterozygous” with respect to the polymorphism. Individuals who are homozygous with respect to one allele are sometimes predisposed to a different phenotype as compared to individuals who are heterozygous or homozygous with respect to another allele.

[0020] In genetic analysis that associate polymorphic variants with osteoarthritis, samples from individuals having osteoarthritis and individuals not having osteoarthritis often are allelotyped and/or genotyped. The term “allelotype” as used herein refers to a process for determining the allele frequency for a polymorphic variant in pooled DNA samples from cases and controls. By pooling DNA from each group, an allele frequency for each SNP in each group is calculated. These allele frequencies are then compared to one another. The term “genotyped” as used herein refers to a process for determining a

genotype of one or more individuals, where a “genotype” is a representation of one or more polymorphic variants in a population.

[0021] A genotype or polymorphic variant may be expressed in terms of a “haplotype,” which as used herein refers to two or more polymorphic variants occurring within genomic DNA in a group of individuals within a population. For example, two SNPs may exist within a gene where each SNP position includes a cytosine variation and an adenine variation. Certain individuals in a population may carry one allele (heterozygous) or two alleles (homozygous) having the gene with a cytosine at each SNP position. As the two cytosines corresponding to each SNP in the gene travel together on one or both alleles in these individuals, the individuals can be characterized as having a cytosine/cytosine haplotype with respect to the two SNPs in the gene.

[0022] As used herein, the term “phenotype” refers to a trait which can be compared between individuals, such as presence or absence of a condition, a visually observable difference in appearance between individuals, metabolic variations, physiological variations, variations in the function of biological molecules, and the like. An example of a phenotype is occurrence of osteoarthritis.

[0023] Researchers sometimes report a polymorphic variant in a database without determining whether the variant is represented in a significant fraction of a population. Because a subset of these reported polymorphic variants are not represented in a statistically significant portion of the population, some of them are sequencing errors and/or not biologically relevant. Thus, it is often not known whether a reported polymorphic variant is statistically significant or biologically relevant until the presence of the variant is detected in a population of individuals and the frequency of the variant is determined. Methods for detecting a polymorphic variant in a population are described herein, specifically in Example 2. A polymorphic variant is statistically significant and often biologically relevant if it is represented in 5% or more of a population, sometimes 10% or more, 15% or more, or 20% or more of a population, and often 25% or more, 30% or more, 35% or more, 40% or more, 45% or more, or 50% or more of a population.

[0024] A polymorphic variant may be detected on either or both strands of a double-stranded nucleic acid. Also, a polymorphic variant may be located within an intron or exon of a gene or within a portion of a regulatory region such as a promoter, a 5′ untranslated region (UTR), a 3′ UTR, and in DNA (e.g., genomic DNA (gDNA) and complementary DNA (cDNA)), RNA (e.g., mRNA, tRNA, and rRNA), or a polypeptide. Polymorphic variations may or may not result in detectable differences in gene expression, polypeptide structure, or polypeptide function.

[0025] It was determined that polymorphic variations associated with an increased risk of osteoarthritis existed in a *ADAMTS2* region in SEQ ID NO: 1. In certain embodiments, a polymorphic variant at position rs398829 in the human genome was associated with an increased risk of osteoarthritis, and in a specific embodiment, a guanine at position rs398829 was associated with an increased risk of osteoarthritis.

[0026] Polymorphic variants in and around the *ADAMTS2* locus were tested for association with osteoarthritis. These include polymorphic variants at positions in SEQ ID NO: 1 selected from the group consisting of 210, 3608, 3609, 4318, 5593, 5629, 5639, 5640, 8943, 17968, 19887, 21034, 21085, , 21596, 23379, 23432, 24007, 26121, 26273, 26755, 27411, 27710, 27842, 28379, 29603, 31232, 31504, 32583, 32794, 32840, 33044, 33150, 33218, 33513, 33959, 34486, 36289, 36570, 38247, 38477, 38518, 38529, 38667, 39781, 39856, 39927, 40506, 41869, 42452, 44788, 46059, 46846, 47712, 48796, 49441, 49602, 49723, 50050, 50171, 50477, 50818, 50833, 50881, 50882, 51386, 51534, 52317, 52368, 52970, 53023, 53356, 53882, 54553, 55475, 55530, 55691, 55848, 55879, 56316, 56911, 57320, 57391, 57437, 57478, 57500, 59111, 59333, 59715, 59804, 59851, 59929, 60052, 60240, 60359, 60381, 60456, 60724, 60875, 60968, 60978, 60998, 61557, 62091, 62645, 62943, 63131, 63145, 63406, 63427, 63554, 63661, 64093, 64153, 64409, 64544, 65257, 65626, 65739, 66392, 66720, 69177, 69336, 69636, 69823, 69928, 70547, 70633, 71805, 72181, 72200, 72474, 72567, 72973, 73468, 73889, 75730, 75970, 76114, 76342, 76449, 76465, 76791, 78042, 80758, 80778, 81356, 81576, 81689, 81759, 81950, 82562, 83591, 83700, 83821, 83842, 83923, 83929, 84021, 84175, 84417, 84747, 85746, 86129, 86335, 87315, 87648, 87764, 87770, 88221, 90474, 91148, 91150, 91160, 91733, 91772, 91785, 93140, 93148, 96080, 96157, 96313, 96759, 97026, 97320, 97732, 98713, 99707, 99959, 100009, 100020, 100065, 100086, 101270, 101276, 101371, 101376, 101439, 101820, 102392, 102602, 102604, 102896, 189104, 189134 and 189205.

Polymorphic variants at the following positions in SEQ ID NO: 1 in particular were associated with an increased risk of osteoarthritis: 5640, 33150, 38247, 38529, 46846, 49723, 50050, 63427, 73889, 189104 and rs428901, where specific embodiments are directed to positions 46846, 73889, 189104 and/or rs428901. In particular, the following polymorphic variants in SEQ ID NO: 1 were associated with risk of osteoarthritis: a cytosine at position 5640, a cytosine at position 33150, an adenine at position 38247, a thymine at position 38529, an adenine at position 46846, a cytosine at position 49723, a cytosine at position 50050, a cytosine at position 63427, a guanine at position 73889, a thymine at position 189104, and an adenine at position rs428901.

[0027] Based in part upon analyses summarized in Figure 1, a region with significant association has been identified in a locus associated with osteoarthritis. Any polymorphic variants associated with osteoarthritis in a region of significant association can be utilized for embodiments described herein. For example, polymorphic variants in a region spanning chromosome positions 178746000 to 178751000 (approximately 5,000 nucleotides in length) in a *ADAMTS2* locus have significant association (chromosome positions are within NCBI's Genome build 34).

Additional Polymorphic Variants Associated with Osteoarthritis

[0028] Also provided is a method for identifying polymorphic variants proximal to an incident, founder polymorphic variant associated with osteoarthritis. Thus, featured herein are methods for

identifying a polymorphic variation associated with osteoarthritis that is proximal to an incident polymorphic variation associated with osteoarthritis, which comprises identifying a polymorphic variant proximal to the incident polymorphic variation associated with osteoarthritis, where the incident polymorphic variant is in a *ADAMTS2* nucleotide sequence. The nucleotide sequence often comprises a polynucleotide sequence selected from the group consisting of (a) a polynucleotide sequence of SEQ ID NO: 1-3; (b) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence encoded by a polynucleotide sequence of SEQ ID NO: 1-3; and (c) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3 or a polynucleotide sequence 90% or more identical to the polynucleotide sequence of SEQ ID NO: 1-3. The presence or absence of an association of the proximal polymorphic variant with osteoarthritis then is determined using a known association method, such as a method described in the Examples hereafter. In an embodiment, the incident polymorphic variant is a polymorphic variant associated with osteoarthritis described herein. In another embodiment, the proximal polymorphic variant identified sometimes is a publicly disclosed polymorphic variant, which for example, sometimes is published in a publicly available database. In other embodiments, the polymorphic variant identified is not publicly disclosed and is discovered using a known method, including, but not limited to, sequencing a region surrounding the incident polymorphic variant in a group of nucleic samples. Thus, multiple polymorphic variants proximal to an incident polymorphic variant are associated with osteoarthritis using this method.

[0029] The proximal polymorphic variant often is identified in a region surrounding the incident polymorphic variant. In certain embodiments, this surrounding region is about 50 kb flanking the first polymorphic variant (e.g. about 50 kb 5' of the first polymorphic variant and about 50 kb 3' of the first polymorphic variant), and the region sometimes is composed of shorter flanking sequences, such as flanking sequences of about 40 kb, about 30 kb, about 25 kb, about 20 kb, about 15 kb, about 10 kb, about 7 kb, about 5 kb, or about 2 kb 5' and 3' of the incident polymorphic variant. In other embodiments, the region is composed of longer flanking sequences, such as flanking sequences of about 55 kb, about 60 kb, about 65 kb, about 70 kb, about 75 kb, about 80 kb, about 85 kb, about 90 kb, about 95 kb, or about 100 kb 5' and 3' of the incident polymorphic variant.

[0030] In certain embodiments, polymorphic variants associated with osteoarthritis are identified iteratively. For example, a first proximal polymorphic variant is associated with osteoarthritis using the methods described above and then another polymorphic variant proximal to the first proximal polymorphic variant is identified (e.g., publicly disclosed or discovered) and the presence or absence of an association of one or more other polymorphic variants proximal to the first proximal polymorphic variant with osteoarthritis is determined.

[0031] The methods described herein are useful for identifying or discovering additional polymorphic variants that may be used to further characterize a gene, region or loci associated with a condition, a disease (e.g., osteoarthritis), or a disorder. For example, allelotyping or genotyping data from the additional polymorphic variants may be used to identify a functional mutation or a region of linkage disequilibrium. In certain embodiments, polymorphic variants identified or discovered within a region comprising the first polymorphic variant associated with osteoarthritis are genotyped using the genetic methods and sample selection techniques described herein, and it can be determined whether those polymorphic variants are in linkage disequilibrium with the first polymorphic variant. The size of the region in linkage disequilibrium with the first polymorphic variant also can be assessed using these genotyping methods. Thus, provided herein are methods for determining whether a polymorphic variant is in linkage disequilibrium with a first polymorphic variant associated with osteoarthritis, and such information can be used in prognosis/diagnosis methods described herein.

Isolated Nucleic Acids

[0032] Featured herein are isolated *ADAMTS2* nucleic acid variants depicted in SEQ ID NO: 1-3, and substantially identical nucleic acids thereof. A nucleic acid variant may be represented on one or both strands in a double-stranded nucleic acid or on one chromosomal complement (heterozygous) or both chromosomal complements (homozygous). *ADAMTS2* exists in two forms, a "long" form comprising a molecule approximately 130 kDa in length (e.g., SEQ ID NO: 2 for cDNA sequence and SEQ ID NO: 4 for amino acid sequence), and a "short" form comprising a molecule approximately 70 kDa in length (e.g., SEQ ID NO: 3 for cDNA sequence and SEQ ID NO: 5 for amino acid sequence). Provided herein are polynucleotide sequences encoding both the short and long forms of *ADAMTS2*.

[0033] As used herein, the term "nucleic acid" includes DNA molecules (e.g., a complementary DNA (cDNA) and genomic DNA (gDNA)) and RNA molecules (e.g., mRNA, rRNA, siRNA and tRNA) and analogs of DNA or RNA, for example, by use of nucleotide analogs. The nucleic acid molecule can be single-stranded and it is often double-stranded. The term "isolated or purified nucleic acid" refers to nucleic acids that are separated from other nucleic acids present in the natural source of the nucleic acid. For example, with regard to genomic DNA, the term "isolated" includes nucleic acids which are separated from the chromosome with which the genomic DNA is naturally associated. An "isolated" nucleic acid is often free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of 5' and/or 3' nucleotide sequences which flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular

material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. As used herein, the term “gene” refers to a nucleotide sequence that encodes a polypeptide.

[0034] Also included herein are nucleic acid fragments. These fragments often have a nucleotide sequence identical to a nucleotide sequence of SEQ ID NO: 1-3, a nucleotide sequence substantially identical to a nucleotide sequence of SEQ ID NO: 1-3, or a nucleotide sequence that is complementary to the foregoing. The nucleic acid fragment may be identical, substantially identical or homologous to a nucleotide sequence in an exon or an intron in a nucleotide sequence of SEQ ID NO: 1-3, and may encode a domain or part of a domain of a polypeptide. Sometimes, the fragment will comprises one or more of the polymorphic variations described herein as being associated with osteoarthritis. The nucleic acid fragment is often 50, 100, or 200 or fewer base pairs in length, and is sometimes about 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 2000, 3000, 4000, 5000, 10000, 15000, or 20000 base pairs in length. A nucleic acid fragment that is complementary to a nucleotide sequence identical or substantially identical to a nucleotide sequence in SEQ ID NO: 1-3 and hybridizes to such a nucleotide sequence under stringent conditions is often referred to as a “probe.” Nucleic acid fragments often include one or more polymorphic sites, or sometimes have an end that is adjacent to a polymorphic site as described hereafter.

[0035] An example of a nucleic acid fragment is an oligonucleotide. As used herein, the term “oligonucleotide” refers to a nucleic acid comprising about 8 to about 50 covalently linked nucleotides, often comprising from about 8 to about 35 nucleotides, and more often from about 10 to about 25 nucleotides. The backbone and nucleotides within an oligonucleotide may be the same as those of naturally occurring nucleic acids, or analogs or derivatives of naturally occurring nucleic acids, provided that oligonucleotides having such analogs or derivatives retain the ability to hybridize specifically to a nucleic acid comprising a targeted polymorphism. Oligonucleotides described herein may be used as hybridization probes or as components of prognostic or diagnostic assays, for example, as described herein.

[0036] Oligonucleotides are typically synthesized using standard methods and equipment, such as the ABIPTM3900 High Throughput DNA Synthesizer and the EXPEDITETM 8909 Nucleic Acid Synthesizer, both of which are available from Applied Biosystems (Foster City, CA). Analogs and derivatives are exemplified in U.S. Pat. Nos. 4,469,863; 5,536,821; 5,541,306; 5,637,683; 5,637,684; 5,700,922; 5,717,083; 5,719,262; 5,739,308; 5,773,601; 5,886,165; 5,929,226; 5,977,296; 6,140,482; WO 00/56746; WO 01/14398, and related publications. Methods for synthesizing oligonucleotides comprising such analogs or derivatives are disclosed, for example, in the patent publications cited above

and in U.S. Pat. Nos. 5,614,622; 5,739,314; 5,955,599; 5,962,674; 6,117,992; in WO 00/75372; and in related publications.

[0037] Oligonucleotides may also be linked to a second moiety. The second moiety may be an additional nucleotide sequence such as a tail sequence (e.g., a polyadenosine tail), an adapter sequence (e.g., phage M13 universal tail sequence), and others. Alternatively, the second moiety may be a non-nucleotide moiety such as a moiety which facilitates linkage to a solid support or a label to facilitate detection of the oligonucleotide. Such labels include, without limitation, a radioactive label, a fluorescent label, a chemiluminescent label, a paramagnetic label, and the like. The second moiety may be attached to any position of the oligonucleotide, provided the oligonucleotide can hybridize to the nucleic acid comprising the polymorphism.

Uses for Nucleic Acid Sequence

[0038] Nucleic acid coding sequences may be used for diagnostic purposes for detection and control of polypeptide expression. Also, included herein are oligonucleotide sequences such as antisense RNA, small-interfering RNA (siRNA) and DNA molecules and ribozymes that function to inhibit translation of a polypeptide. Antisense techniques and RNA interference techniques are known in the art and are described herein.

[0039] Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, hammerhead motif ribozyme molecules may be engineered that specifically and efficiently catalyze endonucleolytic cleavage of RNA sequences corresponding to or complementary to *ADAMTS2* nucleotide sequences. Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between fifteen (15) and twenty (20) ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features such as secondary structure that may render the oligonucleotide sequence unsuitable. The suitability of candidate targets may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

[0040] Antisense RNA and DNA molecules, siRNA and ribozymes may be prepared by any method known in the art for the synthesis of RNA molecules. These include techniques for chemically synthesizing oligodeoxyribonucleotides well known in the art such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences may be incorporated

into a wide variety of vectors which incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

[0041] DNA encoding a polypeptide also may have a number of uses for the diagnosis of diseases, including osteoarthritis, resulting from aberrant expression of a target gene described herein. For example, the nucleic acid sequence may be used in hybridization assays of biopsies or autopsies to diagnose abnormalities of expression or function (e.g., Southern or Northern blot analysis, in situ hybridization assays).

[0042] In addition, the expression of a polypeptide during embryonic development may also be determined using nucleic acid encoding the polypeptide. As addressed, *infra*, production of functionally impaired polypeptide is the cause of various disease states, such as osteoarthritis. *In situ* hybridizations using polypeptide as a probe may be employed to predict problems related to osteoarthritis. Further, as indicated, *infra*, administration of human active polypeptide, recombinantly produced as described herein, may be used to treat disease states related to functionally impaired polypeptide. Alternatively, gene therapy approaches may be employed to remedy deficiencies of functional polypeptide or to replace or compete with dysfunctional polypeptide.

Expression Vectors, Host Cells, and Genetically Engineered Cells

[0043] Provided herein are nucleic acid vectors, often expression vectors, which contain a *ADAMTS2* nucleotide sequence, or a substantially identical sequence thereof. As used herein, the term “vector” refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid, or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors may include replication defective retroviruses, adenoviruses and adeno-associated viruses for example.

[0044] A vector can include a *ADAMTS2* nucleotide sequence in a form suitable for expression of an encoded target polypeptide or target nucleic acid in a host cell. A “target polypeptide” is a polypeptide encoded by a *ADAMTS2* nucleotide sequence, or a substantially identical nucleotide sequence thereof. The recombinant expression vector typically includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term “regulatory sequence” includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of polypeptide desired, and the like. Expression vectors can be introduced into host cells to produce target polypeptides, including fusion polypeptides.

[0045] Recombinant expression vectors can be designed for expression of target polypeptides in prokaryotic or eukaryotic cells. For example, target polypeptides can be expressed in *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[0046] Expression of polypeptides in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion polypeptides. Fusion vectors add a number of amino acids to a polypeptide encoded therein, usually to the amino terminus of the recombinant polypeptide. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant polypeptide; 2) to increase the solubility of the recombinant polypeptide; and 3) to aid in the purification of the recombinant polypeptide by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant polypeptide to enable separation of the recombinant polypeptide from the fusion moiety subsequent to purification of the fusion polypeptide. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith & Johnson, *Gene* 67: 31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding polypeptide, or polypeptide A, respectively, to the target recombinant polypeptide.

[0047] Purified fusion polypeptides can be used in screening assays and to generate antibodies specific for target polypeptides. In a therapeutic embodiment, fusion polypeptide expressed in a retroviral expression vector is used to infect bone marrow cells that are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six (6) weeks).

[0048] Expressing the polypeptide in host bacteria with an impaired capacity to proteolytically cleave the recombinant polypeptide is often used to maximize recombinant polypeptide expression (Gottesman, S., *Gene Expression Technology: Methods in Enzymology*, Academic Press, San Diego, California 185: 119-128 (1990)). Another strategy is to alter the nucleotide sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, *Nucleic Acids Res.* 20: 2111-2118 (1992)). Such alteration of nucleotide sequences can be carried out by standard DNA synthesis techniques.

[0049] When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. Recombinant mammalian expression vectors are

often capable of directing expression of the nucleic acid in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include an albumin promoter (liver-specific; Pinkert *et al.*, *Genes Dev.* 1: 268-277 (1987)), lymphoid-specific promoters (Calame & Eaton, *Adv. Immunol.* 43: 235-275 (1988)), promoters of T cell receptors (Winoto & Baltimore, *EMBO J.* 8: 729-733 (1989)) promoters of immunoglobulins (Banerji *et al.*, *Cell* 33: 729-740 (1983); Queen & Baltimore, *Cell* 33: 741-748 (1983)), neuron-specific promoters (e.g., the neurofilament promoter; Byrne & Ruddle, *Proc. Natl. Acad. Sci. USA* 86: 5473-5477 (1989)), pancreas-specific promoters (Edlund *et al.*, *Science* 230: 912-916 (1985)), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are sometimes utilized, for example, the murine hox promoters (Kessel & Gruss, *Science* 249: 374-379 (1990)) and the α -fetopolypeptide promoter (Carnes & Tilghman, *Genes Dev.* 3: 537-546 (1989)).

[0050] A *ADAMTS2* nucleic acid also may be cloned into an expression vector in an antisense orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a *ADAMTS2* nucleic acid cloned in the antisense orientation can be chosen for directing constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. Antisense expression vectors can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of the regulation of gene expression using antisense genes see, e.g., Weintraub *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) (1986).

[0051] Also provided herein are host cells that include a *ADAMTS2* nucleotide sequence within a recombinant expression vector or a fragment of such a nucleotide sequence which facilitate homologous recombination into a specific site of the host cell genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but rather also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein. A host cell can be any prokaryotic or eukaryotic cell. For example, a target polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[0052] Vectors can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, transduction/infection, DEAE-dextran-mediated transfection, lipofection, or electroporation.

[0053] A host cell provided herein can be used to produce (*i.e.*, express) a target polypeptide or a substantially identical polypeptide thereof. Accordingly, further provided are methods for producing a target polypeptide using host cells described herein. In one embodiment, the method includes culturing host cells into which a recombinant expression vector encoding a target polypeptide has been introduced in a suitable medium such that a target polypeptide is produced. In another embodiment, the method further includes isolating a target polypeptide from the medium or the host cell.

[0054] Also provided are cells or purified preparations of cells which include a *ADAMTS2* transgene, or which otherwise misexpress target polypeptide. Cell preparations can consist of human or non-human cells, *e.g.*, rodent cells, *e.g.*, mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a *ADAMTS2* transgene (*e.g.*, a heterologous form of a *ADAMTS2* gene, such as a human gene expressed in non-human cells). The transgene can be misexpressed, *e.g.*, overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene which misexpress an endogenous target polypeptide (*e.g.*, expression of a gene is disrupted, also known as a knockout). Such cells can serve as a model for studying disorders which are related to mutated or mis-expressed alleles or for use in drug screening. Also provided are human cells (*e.g.*, a hematopoietic stem cells) transfected with a *ADAMTS2* nucleic acid.

[0055] Also provided are cells or a purified preparation thereof (*e.g.*, human cells) in which an endogenous *ADAMTS2* nucleic acid is under the control of a regulatory sequence that does not normally control the expression of the endogenous gene. The expression characteristics of an endogenous gene within a cell (*e.g.*, a cell line or microorganism) can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the corresponding endogenous gene. For example, an endogenous corresponding gene (*e.g.*, a gene which is "transcriptionally silent," not normally expressed, or expressed only at very low levels) may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, *e.g.*, Chappel, US 5,272,071; WO 91/06667, published on May 16, 1991.

Transgenic Animals

[0056] Non-human transgenic animals that express a heterologous target polypeptide (*e.g.*, expressed from a *ADAMTS2* nucleic acid or substantially identical sequence thereof) can be generated. Such animals are useful for studying the function and/or activity of a target polypeptide and for identifying and/or evaluating modulators of the activity of *ADAMTS2* nucleic acids and encoded polypeptides. As used herein, a "transgenic animal" is a non-human animal such as a mammal (*e.g.*, a non-human primate such as chimpanzee, baboon, or macaque; an ungulate such as an equine, bovine, or

caprine; or a rodent such as a rat, a mouse, or an Israeli sand rat), a bird (e.g., a chicken or a turkey), an amphibian (e.g., a frog, salamander, or newt), or an insect (e.g., *Drosophila melanogaster*), in which one or more of the cells of the animal includes a transgene. A transgene is exogenous DNA or a rearrangement (e.g., a deletion of endogenous chromosomal DNA) that is often integrated into or occurs in the genome of cells in a transgenic animal. A transgene can direct expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, and other transgenes can reduce expression (e.g., a knockout). Thus, a transgenic animal can be one in which an endogenous nucleic acid homologous to a *ADAMTS2* nucleic acid has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal (e.g., an embryonic cell of the animal) prior to development of the animal.

[0057] Intronic sequences and polyadenylation signals can also be included in the transgene to increase expression efficiency of the transgene. One or more tissue-specific regulatory sequences can be operably linked to a *ADAMTS2* nucleotide sequence to direct expression of an encoded polypeptide to particular cells. A transgenic founder animal can be identified based upon the presence of a *ADAMTS2* nucleotide sequence in its genome and/or expression of encoded mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a *ADAMTS2* nucleotide sequence can further be bred to other transgenic animals carrying other transgenes.

[0058] Target polypeptides can be expressed in transgenic animals or plants by introducing, for example, a *ADAMTS2* nucleic acid into the genome of an animal that encodes the target polypeptide. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, e.g., a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Also included is a population of cells from a transgenic animal.

Target Polypeptides

[0059] Also featured herein are isolated target polypeptides, which are encoded by a *ADAMTS2* nucleotide sequence (e.g., SEQ ID NO: 1-3), or a substantially identical nucleotide sequence thereof. *ADAMTS2* exists in two forms, a "long" form comprising a molecule approximately 130 kDa in length (e.g., SEQ ID NO: 4), and a "short" form comprising a molecule approximately 70 kDa in length (e.g., SEQ ID NO: 5). Thus featured herein are isolated *ADAMTS2* polypeptides, which include long and short isoforms, and substantially identical polypeptides thereof. An *ADAMTS2* polypeptide is a polypeptide encoded by an *ADAMTS2* nucleic acid, where one nucleic acid can encode one or more different polypeptides. The term "polypeptide" as used herein includes proteins and peptides. An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical

precursors or other chemicals when chemically synthesized. In one embodiment, the language “substantially free” means preparation of a target polypeptide having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-target polypeptide (also referred to herein as a “contaminating protein”), or of chemical precursors or non-target chemicals. When the target polypeptide or a biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, specifically, where culture medium represents less than about 20%, sometimes less than about 10%, and often less than about 5% of the volume of the polypeptide preparation. Isolated or purified target polypeptide preparations are sometimes 0.01 milligrams or more or 0.1 milligrams or more, and often 1.0 milligrams or more and 10 milligrams or more in dry weight.

[0060] Further included herein are target polypeptide fragments. The polypeptide fragment may be a domain or part of a domain of a target polypeptide. The polypeptide fragment may have increased, decreased or unexpected biological activity. The polypeptide fragment is often 50 or fewer, 100 or fewer, or 200 or fewer amino acids in length, and is sometimes 300, 400, 500, 600, 700, or 900 or fewer amino acids in length. Thus, featured herein are *ADAMTS2* polypeptides and biologically active or antigenic fragments thereof useful as reagents or targets in assays applicable to treatment and diagnosis of osteoarthritis. In another embodiment, provided herein are *ADAMTS2* polypeptides having a *ADAMTS2* activity (e.g., a zinc binding activity, a metalloprotease activity, a procollagen II processing or synthesis activity, or a collagen II synthesis activity in vitro or in vivo). In certain embodiments, the polypeptides are *ADAMTS2* proteins including at least one propeptide domain, at least one metalloproteinase domain, at least one disintegrin-like domain, at least one, two, three, and often four thrombospondin domains, and sometimes having a *ADAMTS2* activity, e.g., a *ADAMTS2* activity as described herein. *ADAMTS2* polypeptides and fragments thereof often have biological activity, such as excising the N-propeptide of type II procollagens. Methods for monitoring and quantifying this biological activity are known (e.g., Colige et al., J. Biol. Chem. 270: 16724-16730 (1995)).

[0061] Human *ADAMTS2* protein (SEQ ID NO: 4-5) includes a signal sequence of about 29 amino acids (from amino acid 1 to about amino acid 29 of SEQ ID NO: 4-5). The *ADAMTS2* protein without the signal sequence can be approximately 1182 amino acid residues in length (from about amino acid 30 to amino acid 1211 of SEQ ID NO: 4) or approximately 485 amino acid residues in length (from about amino acid 30 to amino acid 514 of SEQ ID NO: 5). Human *ADAMTS2* protein includes a “pro” region homologous to the repolysin family propeptide, which is typically post-translationally cleaved upon conversion of the inactive (or pro-domain containing) protein to the catalytically active metalloprotease. The prodomain region of human *ADAMTS2* protein corresponds to about amino acids 30 to 251, 30 to 252, 30 to 253, 30 to 254, 30 to 255, 30 to 256, 30 to 257, 30 to 258 or 30 to 259 of SEQ ID NO: 4-5, where it is understood that the active form of *ADAMTS2* does not contain the propeptide domain.

[0062] Upon cleavage, catalytically active mature protein can be approximately 960, 959, 958, 957, 956, 955, 954, 953 or 952 amino acids in length (from about amino acid 252, 253, 254, 255, 256, 257, 258, 259 or 260 to amino acid 1211 of SEQ ID NO: 4) or approximately 261, 260, 259, 258, 257, 256, 255, 254 or 253 amino acid residues in length (from about amino acid 252, 253, 254, 255, 256, 257, 258, 259 or 260 to amino acid 514 of SEQ ID NO: 5).

[0063] Human *ADAMTS2* contains the following regions or other structural features: a signal sequence at about amino acids 1-29 of SEQ ID NO: 4-5; a reprolysin family propeptide domain located at about amino acid residues 30 to 251, 30 to 252, 30 to 253, 30 to 254, 30 to 255, 30 to 256, 30 to 257, 30 to 258 or 30 to 259 of SEQ ID NO: 4-5; a zinc-metalloprotease catalytic domain at about amino acids 251 to 479, 252 to 479, 253 to 479, 254 to 479, 255 to 479, 256 to 479, 257 to 479, 258 to 479 or 259 to 479 of SEQ ID NO: 4-5; a disintegrin domain at about amino acids 480 to 560 of SEQ ID NO: 4; a cysteine-rich domain at about amino acids 618 to 722 of SEQ ID NO: 4; four thrombospondin motifs-2 motifs at about amino acids 561 to 616, 854 to 912, 914 to 971, and 975 to 1029 of SEQ ID NO: 4; and eight N-glycosylation sites located at about amino acids 112, 251, 949, 993, 1031, 1098, 1145, and 1150 of SEQ ID NO: 4.

[0064] In other embodiments, provided are methods of increasing the synthesis of procollagen II comprising providing or administering to individuals in need of increasing levels of type II collagen the pharmaceutical or physiologically acceptable composition comprising active human *ADAMTS2* protein or fragment thereof, where *ADAMTS2* polypeptide fragments having activity are selected from amino acids 252-1211, 253-1211, 254-1211, 255-1211, 256-1211, 257-1211, 258-1211, 259-1211 or 260-1211 of SEQ ID NO: 4, where it is understood that the active form of *ADAMTS2* does not contain the propeptide domain.

[0065] Substantially identical target polypeptides may depart from the amino acid sequences of target polypeptides in different manners. For example, conservative amino acid modifications may be introduced at one or more positions in the amino acid sequences of target polypeptides. A "conservative amino acid substitution" is one in which the amino acid is replaced by another amino acid having a similar structure and/or chemical function. Families of amino acid residues having similar structures and functions are well known. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Also, essential and non-essential amino acids may be replaced. A "non-essential" amino acid is one that can be altered without abolishing or substantially altering the biological function of a target polypeptide, whereas altering an "essential" amino acid abolishes or substantially alters the

biological function of a target polypeptide. Amino acids that are conserved among target polypeptides are typically essential amino acids. In certain embodiments, the polypeptide includes one or more non-synonymous polymorphic variants associated with osteoarthritis.

[0066] Also, target polypeptides may exist as chimeric or fusion polypeptides. As used herein, a target “chimeric polypeptide” or target “fusion polypeptide” includes a target polypeptide linked to a non-target polypeptide. A “non-target polypeptide” refers to a polypeptide having an amino acid sequence corresponding to a polypeptide which is not substantially identical to the target polypeptide, which includes, for example, a polypeptide that is different from the target polypeptide and derived from the same or a different organism. The target polypeptide in the fusion polypeptide can correspond to an entire or nearly entire target polypeptide or a fragment thereof. The non-target polypeptide can be fused to the N-terminus or C-terminus of the target polypeptide.

[0067] Fusion polypeptides can include a moiety having high affinity for a ligand. For example, the fusion polypeptide can be a GST-target fusion polypeptide in which the target sequences are fused to the C-terminus of the GST sequences, or a polyhistidine-target fusion polypeptide in which the target polypeptide is fused at the N- or C-terminus to a string of histidine residues. Such fusion polypeptides can facilitate purification of recombinant target polypeptide. Expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide), and a nucleotide sequence in SEQ ID NO: 1-3, or a substantially identical nucleotide sequence thereof, can be cloned into an expression vector such that the fusion moiety is linked in-frame to the target polypeptide. Further, the fusion polypeptide can be a target polypeptide containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression, secretion, cellular internalization, and cellular localization of a target polypeptide can be increased through use of a heterologous signal sequence. Fusion polypeptides can also include all or a part of a serum polypeptide (e.g., an IgG constant region or human serum albumin).

[0068] Target polypeptides can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. Administration of these target polypeptides can be used to affect the bioavailability of a substrate of the target polypeptide and may effectively increase target polypeptide biological activity in a cell. Target fusion polypeptides may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a target polypeptide; (ii) misregulation of the gene encoding the target polypeptide; and (iii) aberrant post-translational modification of a target polypeptide. Also, target polypeptides can be used as immunogens to produce anti-target antibodies in a subject, to purify target polypeptide ligands or binding partners, and in screening assays to identify molecules which inhibit or enhance the interaction of a target polypeptide with a substrate.

[0069] In addition, polypeptides can be chemically synthesized using techniques known in the art (See, e.g., Creighton, 1983 *Proteins*. New York, N.Y.: W. H. Freeman and Company; and Hunkapiller et

al., (1984) Nature July 12 -18;310(5973):105-11). For example, a relative short fragment can be synthesized by use of a peptide synthesizer. Furthermore, if desired, non-classical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the fragment sequence. Non-classical amino acids include, but are not limited to, to the D-isomers of the common amino acids, 2,4-diaminobutyric acid, α-amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid, g-Abu, ε-Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, homocitrulline, cysteine acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, β-alanine, fluoroamino acids, designer amino acids such as β-methyl amino acids, Ca-methyl amino acids, Na-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

[0070] Polypeptides and polypeptide fragments sometimes are differentially modified during or after translation, e.g., by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including but not limited, to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH₄; acetylation, formylation, oxidation, reduction; metabolic synthesis in the presence of tunicamycin; and the like. Additional post-translational modifications include, for example, N-linked or O-linked carbohydrate chains, processing of N-terminal or C-terminal ends), attachment of chemical moieties to the amino acid backbone, chemical modifications of N-linked or O-linked carbohydrate chains, and addition or deletion of an N-terminal methionine residue as a result of prokaryotic host cell expression. The polypeptide fragments may also be modified with a detectable label, such as an enzymatic, fluorescent, isotopic or affinity label to allow for detection and isolation of the polypeptide.

[0071] Also provided are chemically modified derivatives of polypeptides that can provide additional advantages such as increased solubility, stability and circulating time of the polypeptide, or decreased immunogenicity (see e.g., U.S. Pat. No: 4,179,337. The chemical moieties for derivitization may be selected from water soluble polymers such as polyethylene glycol, ethylene glycol/propylene glycol copolymers, carboxymethylcellulose, dextran, polyvinyl alcohol and the like. The polypeptides may be modified at random positions within the molecule, or at predetermined positions within the molecule and may include one, two, three or more attached chemical moieties.

[0072] The polymer may be of any molecular weight, and may be branched or unbranched. For polyethylene glycol, the preferred molecular weight is between about 1 kDa and about 100 kDa (the term "about" indicating that in preparations of polyethylene glycol, some molecules will weigh more, some less, than the stated molecular weight) for ease in handling and manufacturing. Other sizes may be used, depending on the desired therapeutic profile (e.g., the duration of sustained release desired, the effects, if

any on biological activity, the ease in handling, the degree or lack of antigenicity and other known effects of the polyethylene glycol to a therapeutic protein or analog).

[0073] The polymers should be attached to the polypeptide with consideration of effects on functional or antigenic domains of the polypeptide. There are a number of attachment methods available to those skilled in the art (e.g., EP 0 401 384 (coupling PEG to G-CSF) and Malik et al. (1992) Exp Hematol. September;20(8):1028-35 (pegylation of GM-CSF using tresyl chloride)). For example, polyethylene glycol may be covalently bound through amino acid residues via a reactive group, such as a free amino or carboxyl group. Reactive groups are those to which an activated polyethylene glycol molecule may be bound. The amino acid residues having a free amino group may include lysine residues and the N-terminal amino acid residues; those having a free carboxyl group may include aspartic acid residues, glutamic acid residues and the C-terminal amino acid residue. Sulfhydryl groups may also be used as a reactive group for attaching the polyethylene glycol molecules. For therapeutic purposes, the attachment sometimes is at an amino group, such as attachment at the N-terminus or lysine group.

[0074] Proteins can be chemically modified at the N-terminus. Using polyethylene glycol as an illustration of such a composition, one may select from a variety of polyethylene glycol molecules (by molecular weight, branching, and the like), the proportion of polyethylene glycol molecules to protein (polypeptide) molecules in the reaction mix, the type of pegylation reaction to be performed, and the method of obtaining the selected N-terminally pegylated protein. The method of obtaining the N-terminally pegylated preparation (i.e., separating this moiety from other monopegylated moieties if necessary) may be by purification of the N-terminally pegylated material from a population of pegylated protein molecules. Selective proteins chemically modified at the N-terminus may be accomplished by reductive alkylation, which exploits differential reactivity of different types of primary amino groups (lysine versus the N-terminal) available for derivatization in a particular protein. Under the appropriate reaction conditions, substantially selective derivatization of the protein at the N-terminus with a carbonyl group containing polymer is achieved.

Substantially Identical Nucleic Acids and Polypeptides

[0075] Nucleotide sequences and polypeptide sequences that are substantially identical to a *ADAMTS2* nucleotide sequence and the target polypeptide sequences encoded by those nucleotide sequences, respectively, are included herein. The term "substantially identical" as used herein refers to two or more nucleic acids or polypeptides sharing one or more identical nucleotide sequences or polypeptide sequences, respectively. Included are nucleotide sequences or polypeptide sequences that are 55% or more, 60% or more, 65% or more, 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, 95% or more (each often within a 1%, 2%, 3% or 4% variability) identical to a *ADAMTS2* nucleotide sequence or the encoded target polypeptide amino acid sequences. One test for determining

whether two nucleic acids are substantially identical is to determine the percent of identical nucleotide sequences or polypeptide sequences shared between the nucleic acids or polypeptides.

[0076] Calculations of sequence identity are often performed as follows. Sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is sometimes 30% or more, 40% or more, 50% or more, often 60% or more, and more often 70% or more, 80% or more, 90% or more, or 100% of the length of the reference sequence. The nucleotides or amino acids at corresponding nucleotide or polypeptide positions, respectively, are then compared among the two sequences. When a position in the first sequence is occupied by the same nucleotide or amino acid as the corresponding position in the second sequence, the nucleotides or amino acids are deemed to be identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, introduced for optimal alignment of the two sequences.

[0077] Comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. Percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of Meyers & Miller, *CABIOS* 4: 11-17 (1989), which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. Also, percent identity between two amino acid sequences can be determined using the Needleman & Wunsch, *J. Mol. Biol.* 48: 444-453 (1970) algorithm which has been incorporated into the GAP program in the GCG software package (available at the [http address www.gcg.com](http://www.gcg.com)), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. Percent identity between two nucleotide sequences can be determined using the GAP program in the GCG software package (available at [http address www.gcg.com](http://www.gcg.com)), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A set of parameters often used is a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[0078] Another manner for determining if two nucleic acids are substantially identical is to assess whether a polynucleotide homologous to one nucleic acid will hybridize to the other nucleic acid under stringent conditions. As use herein, the term "stringent conditions" refers to conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989). Aqueous and non-aqueous methods are described in that reference and either can be used. An example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C,

followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Often, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. More often, stringency conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C.

[0079] An example of a substantially identical nucleotide sequence to a nucleotide sequence in SEQ ID NO: 1-3 is one that has a different nucleotide sequence but still encodes the same polypeptide sequence encoded by the nucleotide sequence in SEQ ID NO: 1-3. Another example is a nucleotide sequence that encodes a polypeptide having a polypeptide sequence that is more than 70% or more identical to, sometimes more than 75% or more, 80% or more, or 85% or more identical to, and often more than 90% or more and 95% or more identical to a polypeptide sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3.

[0080] Nucleotide sequences in SEQ ID NO: 1-3 and amino acid sequences of encoded polypeptides can be used as “query sequences” to perform a search against public databases to identify other family members or related sequences, for example. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.*, *J. Mol. Biol.* 215: 403-10 (1990). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to nucleotide sequences in SEQ ID NO: 1-3. BLAST polypeptide searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to polypeptides encoded by the nucleotide sequences of SEQ ID NO: 1-3. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, *Nucleic Acids Res.* 25(17): 3389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used (*see* the http address www.ncbi.nlm.nih.gov).

[0081] A nucleic acid that is substantially identical to a nucleotide sequence in SEQ ID NO: 1-3 may include polymorphic sites at positions equivalent to those described herein when the sequences are aligned. For example, using the alignment procedures described herein, SNPs in a sequence substantially identical to a sequence in SEQ ID NO: 1-3 can be identified at nucleotide positions that match (*i.e.*, align) with nucleotides at SNP positions in each nucleotide sequence in SEQ ID NO: 1-3. Also, where a polymorphic variation results in an insertion or deletion, insertion or deletion of a nucleotide sequence

from a reference sequence can change the relative positions of other polymorphic sites in the nucleotide sequence.

[0082] Substantially identical nucleotide and polypeptide sequences include those that are naturally occurring, such as allelic variants (same locus), splice variants, homologs (different locus), and orthologs (different organism) or can be non-naturally occurring. Non-naturally occurring variants can be generated by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product). Orthologs, homologs, allelic variants, and splice variants can be identified using methods known in the art. These variants normally comprise a nucleotide sequence encoding a polypeptide that is 50% or more, about 55% or more, often about 70-75% or more or about 80-85% or more, and sometimes about 90-95% or more identical to the amino acid sequences of target polypeptides or a fragment thereof. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions to a nucleotide sequence in SEQ ID NO: 1-3 or a fragment of this sequence. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of a nucleotide sequence in SEQ ID NO: 1-3 can further be identified by mapping the sequence to the same chromosome or locus as the nucleotide sequence in SEQ ID NO: 1-3.

[0083] Also, substantially identical nucleotide sequences may include codons that are altered with respect to the naturally occurring sequence for enhancing expression of a target polypeptide in a particular expression system. For example, the nucleic acid can be one in which one or more codons are altered, and often 10% or more or 20% or more of the codons are altered for optimized expression in bacteria (e.g., *E. coli*), yeast (e.g., *S. cerevisiae*), human (e.g., 293 cells), insect, or rodent (e.g., hamster) cells.

Methods for Identifying Risk of Osteoarthritis

[0084] Methods for prognosing and diagnosing osteoarthritis are included herein. These methods include detecting the presence or absence of one or more polymorphic variations in a nucleotide sequence associated with osteoarthritis, such as variants in or around the loci set forth herein, or a substantially identical sequence thereof, in a sample from a subject, where the presence of a polymorphic variant described herein is indicative of a risk of osteoarthritis. Determining a risk of osteoarthritis sometimes refers to determining whether an individual is at an increased risk of osteoarthritis (e.g., intermediate risk or higher risk).

[0085] Thus, featured herein is a method for identifying a subject who is at risk of osteoarthritis, which comprises detecting an aberration associated with osteoarthritis in a nucleic acid sample from the

subject. An embodiment is a method for detecting a risk of osteoarthritis in a subject, which comprises detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject, where the nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of: (a) a nucleotide sequence of SEQ ID NO: 1-3; (b) a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3; (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-3; and (d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising the polymorphic site; whereby the presence of the polymorphic variation is indicative of a predisposition to osteoarthritis in the subject. In some embodiments, a polymorphic variation at position 733 of SEQ ID NO: 2-3 may be detected (e.g., a guanine in these sequences or a cytosine in a complementary sequence are associated with increased risk of osteoarthritis). In certain embodiments, polymorphic variants at the positions described herein are detected for determining a risk of osteoarthritis, and polymorphic variants at positions in linkage disequilibrium with these positions are detected for determining a risk of osteoarthritis. As used herein, "SEQ ID NO: 1-3" refers to individual sequences in SEQ ID NO: 1, 2 or 3, each sequence being separately applicable to embodiments described herein.

[0086] Risk of osteoarthritis sometimes is expressed as a probability, such as an odds ratio, percentage, or risk factor. Risk often is based upon the presence or absence of one or more polymorphic variants described herein, and also may be based in part upon phenotypic traits of the individual being tested. Methods for calculating risk based upon patient data are well known (*see, e.g., Agresti, Categorical Data Analysis*, 2nd Ed. 2002. Wiley). Allelotyping and genotyping analyses may be carried out in populations other than those exemplified herein to enhance the predictive power of the prognostic method. These further analyses are executed in view of the exemplified procedures described herein, and may be based upon the same polymorphic variations or additional polymorphic variations.

[0087] In certain embodiments, determining the presence of a combination of two or more polymorphic variants associated with osteoarthritis in one or more genetic loci (e.g., one or more genes) of the sample is determined to identify, quantify and/or estimate, risk of osteoarthritis. The risk often is the probability of having or developing osteoarthritis. The risk sometimes is expressed as a relative risk with respect to a population average risk of osteoarthritis, and sometimes is expressed as a relative risk with respect to the lowest risk group. Such relative risk assessments often are based upon penetrance values determined by statistical methods, and are particularly useful to clinicians and insurance companies for assessing risk of osteoarthritis (e.g., a clinician can target appropriate detection, prevention and therapeutic regimens to a patient after determining the patient's risk of osteoarthritis, and an

insurance company can fine tune actuarial tables based upon population genotype assessments of osteoarthritis risk). Risk of osteoarthritis sometimes is expressed as an odds ratio, which is the odds of a particular person having a genotype has or will develop osteoarthritis with respect to another genotype group (e.g., the most disease protective genotype or population average). In related embodiments, the determination is utilized to identify a subject at risk of osteoarthritis. In an embodiment, two or more polymorphic variations are detected in two or more regions in human genomic DNA associated with increased risk of osteoarthritis, such as a locus containing a *ADAMTS2*, for example. In certain embodiments, 3 or more, or 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 or more polymorphic variants are detected in the sample. In specific embodiments, polymorphic variants are detected in a *ADAMTS2* region, for example. In certain embodiments, polymorphic variants are detected at other genetic loci (e.g., the polymorphic variants can be detected in *ADAMTS2* in addition to other loci or only in other loci), where the other loci include but are not limited to those described in concurrently-filed patent applications having attorney docket number 524593008700, 524593008800, 524593009000 or 524593009200, which is incorporated herein by reference in its entirety.

[0088] Results from prognostic tests may be combined with other test results to diagnose osteoarthritis. For example, prognostic results may be gathered, a patient sample may be ordered based on a determined predisposition to osteoarthritis, the patient sample is analyzed, and the results of the analysis may be utilized to diagnose osteoarthritis. Also osteoarthritis diagnostic method can be developed from studies used to generate prognostic methods in which populations are stratified into subpopulations having different progressions of osteoarthritis. In another embodiment, prognostic results may be gathered, a patient's risk factors for developing osteoarthritis (e.g., age, weight, race, diet) analyzed, and a patient sample may be ordered based on a determined predisposition to osteoarthritis.

[0089] The nucleic acid sample typically is isolated from a biological sample obtained from a subject. For example, nucleic acid can be isolated from blood, saliva, sputum, urine, cell scrapings, and biopsy tissue. The nucleic acid sample can be isolated from a biological sample using standard techniques, such as the technique described in Example 2. As used herein, the term "subject" refers primarily to humans but also refers to other mammals such as dogs, cats, and ungulates (e.g., cattle, sheep, and swine). Subjects also include avians (e.g., chickens and turkeys), reptiles, and fish (e.g., salmon), as embodiments described herein can be adapted to nucleic acid samples isolated from any of these organisms. The nucleic acid sample may be isolated from the subject and then directly utilized in a method for determining the presence of a polymorphic variant, or alternatively, the sample may be isolated and then stored (e.g., frozen) for a period of time before being subjected to analysis.

[0090] The presence or absence of a polymorphic variant is determined using one or both chromosomal complements represented in the nucleic acid sample. Determining the presence or absence of a polymorphic variant in both chromosomal complements represented in a nucleic acid sample from a

subject having a copy of each chromosome is useful for determining the zygosity of an individual for the polymorphic variant (*i.e.*, whether the individual is homozygous or heterozygous for the polymorphic variant). Any oligonucleotide-based diagnostic may be utilized to determine whether a sample includes the presence or absence of a polymorphic variant in a sample. For example, primer extension methods, ligase sequence determination methods (*e.g.*, U.S. Pat. Nos. 5,679,524 and 5,952,174, and WO 01/27326), mismatch sequence determination methods (*e.g.*, U.S. Pat. Nos. 5,851,770; 5,958,692; 6,110,684; and 6,183,958), microarray sequence determination methods, restriction fragment length polymorphism (RFLP), single strand conformation polymorphism detection (SSCP) (*e.g.*, U.S. Pat. Nos. 5,891,625 and 6,013,499), PCR-based assays (*e.g.*, TAQMAN® PCR System (Applied Biosystems)), and nucleotide sequencing methods may be used.

[0091] Oligonucleotide extension methods typically involve providing a pair of oligonucleotide primers in a polymerase chain reaction (PCR) or in other nucleic acid amplification methods for the purpose of amplifying a region from the nucleic acid sample that comprises the polymorphic variation. One oligonucleotide primer is complementary to a region 3' of the polymorphism and the other is complementary to a region 5' of the polymorphism. A PCR primer pair may be used in methods disclosed in U.S. Pat. Nos. 4,683,195; 4,683,202, 4,965,188; 5,656,493; 5,998,143; 6,140,054; WO 01/27327; and WO 01/27329 for example. PCR primer pairs may also be used in any commercially available machines that perform PCR, such as any of the GENEAMP® Systems available from Applied Biosystems. Also, those of ordinary skill in the art will be able to design oligonucleotide primers based upon a *ADAMTS2* nucleotide sequence using knowledge available in the art.

[0092] Also provided is an extension oligonucleotide that hybridizes to the amplified fragment adjacent to the polymorphic variation. As used herein, the term "adjacent" refers to the 3' end of the extension oligonucleotide being often 1 nucleotide from the 5' end of the polymorphic site, and sometimes 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides from the 5' end of the polymorphic site, in the nucleic acid when the extension oligonucleotide is hybridized to the nucleic acid. The extension oligonucleotide then is extended by one or more nucleotides, and the number and/or type of nucleotides that are added to the extension oligonucleotide determine whether the polymorphic variant is present. Oligonucleotide extension methods are disclosed, for example, in U.S. Pat. Nos. 4,656,127; 4,851,331; 5,679,524; 5,834,189; 5,876,934; 5,908,755; 5,912,118; 5,976,802; 5,981,186; 6,004,744; 6,013,431; 6,017,702; 6,046,005; 6,087,095; 6,210,891; and WO 01/20039. Oligonucleotide extension methods using mass spectrometry are described, for example, in U.S. Pat. Nos. 5,547,835; 5,605,798; 5,691,141; 5,849,542; 5,869,242; 5,928,906; 6,043,031; and 6,194,144, and a method often utilized is described herein in Example 2.

[0093] A microarray can be utilized for determining whether a polymorphic variant is present or absent in a nucleic acid sample. A microarray may include any oligonucleotides described herein, and

methods for making and using oligonucleotide microarrays suitable for diagnostic use are disclosed in U.S. Pat. Nos. 5,492,806; 5,525,464; 5,589,330; 5,695,940; 5,849,483; 6,018,041; 6,045,996; 6,136,541; 6,142,681; 6,156,501; 6,197,506; 6,223,127; 6,225,625; 6,229,911; 6,239,273; WO 00/52625; WO 01/25485; and WO 01/29259. The microarray typically comprises a solid support and the oligonucleotides may be linked to this solid support by covalent bonds or by non-covalent interactions. The oligonucleotides may also be linked to the solid support directly or by a spacer molecule. A microarray may comprise one or more oligonucleotides complementary to a polymorphic site set forth herein.

[0094] A kit also may be utilized for determining whether a polymorphic variant is present or absent in a nucleic acid sample. A kit often comprises one or more pairs of oligonucleotide primers useful for amplifying a fragment of a nucleotide sequence of SEQ ID NO: 1-3 or a substantially identical sequence thereof, where the fragment includes a polymorphic site. The kit sometimes comprises a polymerizing agent, for example, a thermostable nucleic acid polymerase such as one disclosed in U.S. Pat. Nos. 4,889,818 or 6,077,664. Also, the kit often comprises an elongation oligonucleotide that hybridizes to a *ADAMTS2* nucleotide sequence in a nucleic acid sample adjacent to the polymorphic site. Where the kit includes an elongation oligonucleotide, it also often comprises chain elongating nucleotides, such as dATP, dTTP, dGTP, dCTP, and dITP, including analogs of dATP, dTTP, dGTP, dCTP and dITP, provided that such analogs are substrates for a thermostable nucleic acid polymerase and can be incorporated into a nucleic acid chain elongated from the extension oligonucleotide. Along with chain elongating nucleotides would be one or more chain terminating nucleotides such as ddATP, ddTTP, ddGTP, ddCTP, and the like. In an embodiment, the kit comprises one or more oligonucleotide primer pairs, a polymerizing agent, chain elongating nucleotides, at least one elongation oligonucleotide, and one or more chain terminating nucleotides. Kits optionally include buffers, vials, microtiter plates, and instructions for use.

[0095] An individual identified as being at risk of osteoarthritis may be heterozygous or homozygous with respect to the allele associated with a higher risk of osteoarthritis. A subject homozygous for an allele associated with an increased risk of osteoarthritis is at a comparatively high risk of osteoarthritis, a subject heterozygous for an allele associated with an increased risk of osteoarthritis is at a comparatively intermediate risk of osteoarthritis, and a subject homozygous for an allele associated with a decreased risk of osteoarthritis is at a comparatively low risk of osteoarthritis. A genotype may be assessed for a complementary strand, such that the complementary nucleotide at a particular position is detected.

[0096] Also featured are methods for determining risk of osteoarthritis and/or identifying a subject at risk of osteoarthritis by contacting a polypeptide or protein encoded by a *ADAMTS2* nucleotide sequence from a subject with an antibody that specifically binds to an epitope associated with increased

risk of osteoarthritis in the polypeptide (e.g., an epitope comprising a valine at position 245 in an *IRILRI* polypeptide).

Applications of Prognostic and Diagnostic Results to Pharmacogenomic Methods

[0097] Pharmacogenomics is a discipline that involves tailoring a treatment for a subject according to the subject's genotype as a particular treatment regimen may exert a differential effect depending upon the subject's genotype. For example, based upon the outcome of a prognostic test described herein, a clinician or physician may target pertinent information and preventative or therapeutic treatments to a subject who would be benefited by the information or treatment and avoid directing such information and treatments to a subject who would not be benefited (e.g., the treatment has no therapeutic effect and/or the subject experiences adverse side effects).

[0098] The following is an example of a pharmacogenomic embodiment. A particular treatment regimen can exert a differential effect depending upon the subject's genotype. Where a candidate therapeutic exhibits a significant interaction with a major allele and a comparatively weak interaction with a minor allele (e.g., an order of magnitude or greater difference in the interaction), such a therapeutic typically would not be administered to a subject genotyped as being homozygous for the minor allele, and sometimes not administered to a subject genotyped as being heterozygous for the minor allele. In another example, where a candidate therapeutic is not significantly toxic when administered to subjects who are homozygous for a major allele but is comparatively toxic when administered to subjects heterozygous or homozygous for a minor allele, the candidate therapeutic is not typically administered to subjects who are genotyped as being heterozygous or homozygous with respect to the minor allele.

[0099] The methods described herein are applicable to pharmacogenomic methods for preventing, alleviating or treating osteoarthritis. For example, a nucleic acid sample from an individual may be subjected to a prognostic test described herein. Where one or more polymorphic variations associated with increased risk of osteoarthritis are identified in a subject, information for preventing or treating osteoarthritis and/or one or more osteoarthritis treatment regimens then may be prescribed to that subject.

[0100] In certain embodiments, a treatment or preventative regimen is specifically prescribed and/or administered to individuals who will most benefit from it based upon their risk of developing osteoarthritis assessed by the methods described herein. Thus, provided are methods for identifying a subject predisposed to osteoarthritis and then prescribing a therapeutic or preventative regimen to individuals identified as having a predisposition. Thus, certain embodiments are directed to a method for reducing osteoarthritis in a subject, which comprises: detecting the presence or absence of a polymorphic variant associated with osteoarthritis in a nucleotide sequence in a nucleic acid sample from a subject, where the nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of: (a) a nucleotide sequence of SEQ ID NO: 1-3; (b) a nucleotide sequence which encodes a polypeptide

consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3; (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-3; and (d) a fragment of a polynucleotide sequence of (a), (b), or (c); and prescribing or administering a treatment regimen to a subject from whom the sample originated where the presence of a polymorphic variation associated with osteoarthritis is detected in the nucleotide sequence. In these methods, predisposition results may be utilized in combination with other test results to diagnose osteoarthritis.

[0101] Certain preventative treatments often are prescribed to subjects having a predisposition to osteoarthritis and where the subject is diagnosed with osteoarthritis or is diagnosed as having symptoms indicative of an early stage of osteoarthritis. The treatment sometimes is preventative (e.g., is prescribed or administered to reduce the probability that osteoarthritis arises or progresses), sometimes is therapeutic, and sometimes delays, alleviates or halts the progression of osteoarthritis. Any known preventative or therapeutic treatment for alleviating or preventing the occurrence of osteoarthritis is prescribed and/or administered. For example, the treatment often is directed to decreasing pain and improving joint movement. Examples of OA treatments include exercises to keep joints flexible and improve muscle strength. Different medications to control pain, including corticosteroids and nonsteroidal anti-inflammatory drugs (NSAIDs, e.g., Voltaren); cyclooxygenase-2 (COX-2) inhibitors (e.g., Celebrex, Vioxx, Mobic, and Bextra); monoclonal antibodies (e.g., Remicade); tumor necrosis factor inhibitors (e.g., Enbrel); or injections of glucocorticoids, hyaluronic acid or chondroitin sulfate into joints that are inflamed and not responsive to NSAIDs. Orally administered chondroitin sulfate also may be used as a therapeutic, as it may increase hyaluronic acid levels and viscosity of synovial fluid, and decrease collagenase levels in synovial fluid. Also, glucosamine can serve as an OA therapeutic as delivering it into joints may inhibit enzymes involved in cartilage degradation and enhance the production of hyaluronic acid. For mild pain without inflammation, acetaminophen may be used. Other treatments include: heat/cold therapy for temporary pain relief; joint protection to prevent strain or stress on painful joints; surgery to relieve chronic pain in damaged joints; and weight control to prevent extra stress on weight-bearing joints.

[0102] As therapeutic approaches for treating osteoarthritis continue to evolve and improve, the goal of treatments for osteoarthritis related disorders is to intervene even before clinical signs first manifest. Thus, genetic markers associated with susceptibility to osteoarthritis prove useful for early diagnosis, prevention and treatment of osteoarthritis.

[0103] As osteoarthritis preventative and treatment information can be specifically targeted to subjects in need thereof (e.g., those at risk of developing osteoarthritis or those in an early stage of osteoarthritis), provided herein is a method for preventing or reducing the risk of developing

osteoarthritis in a subject, which comprises: (a) detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) identifying a subject with a predisposition to osteoarthritis, whereby the presence of the polymorphic variation is indicative of a predisposition to osteoarthritis in the subject; and (c) if such a predisposition is identified, providing the subject with information about methods or products to prevent or reduce osteoarthritis or to delay the onset of osteoarthritis. Also provided is a method of targeting information or advertising to a subpopulation of a human population based on the subpopulation being genetically predisposed to a disease or condition, which comprises: (a) detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) identifying the subpopulation of subjects in which the polymorphic variation is associated with osteoarthritis; and (c) providing information only to the subpopulation of subjects about a particular product which may be obtained and consumed or applied by the subject to help prevent or delay onset of the disease or condition.

[0104] Pharmacogenomics methods also may be used to analyze and predict a response to osteoarthritis treatment or a drug. For example, if pharmacogenomics analysis indicates a likelihood that an individual will respond positively to osteoarthritis treatment with a particular drug, the drug may be administered to the individual. Conversely, if the analysis indicates that an individual is likely to respond negatively to treatment with a particular drug, an alternative course of treatment may be prescribed. A negative response may be defined as either the absence of an efficacious response or the presence of toxic side effects. The response to a therapeutic treatment can be predicted in a background study in which subjects in any of the following populations are genotyped: a population that responds favorably to a treatment regimen, a population that does not respond significantly to a treatment regimen, and a population that responds adversely to a treatment regimen (e.g., exhibits one or more side effects). These populations are provided as examples and other populations and subpopulations may be analyzed. Based upon the results of these analyses, a subject is genotyped to predict whether he or she will respond favorably to a treatment regimen, not respond significantly to a treatment regimen, or respond adversely to a treatment regimen.

[0105] The tests described herein also are applicable to clinical drug trials. One or more polymorphic variants indicative of response to an agent for treating osteoarthritis or to side effects to an agent for treating osteoarthritis may be identified using the methods described herein. Thereafter, potential participants in clinical trials of such an agent may be screened to identify those individuals most likely to respond favorably to the drug and exclude those likely to experience side effects. In that way, the effectiveness of drug treatment may be measured in individuals who respond positively to the drug, without lowering the measurement as a result of the inclusion of individuals who are unlikely to respond positively in the study and without risking undesirable safety problems.

[0106] Thus, another embodiment is a method of selecting an individual for inclusion in a clinical trial of a treatment or drug comprising the steps of: (a) obtaining a nucleic acid sample from an individual; (b) determining the identity of a polymorphic variation which is associated with a positive response to the treatment or the drug, or at least one polymorphic variation which is associated with a negative response to the treatment or the drug in the nucleic acid sample, and (c) including the individual in the clinical trial if the nucleic acid sample contains said polymorphic variation associated with a positive response to the treatment or the drug or if the nucleic acid sample lacks said polymorphic variation associated with a negative response to the treatment or the drug. In addition, the methods described herein for selecting an individual for inclusion in a clinical trial of a treatment or drug encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination. The polymorphic variation may be in a sequence selected individually or in any combination from the group consisting of (i) a nucleotide sequence of SEQ ID NO: 1-3; (ii) a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3; (iii) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-3; and (iv) a fragment of a polynucleotide sequence of (i), (ii), or (iii) comprising the polymorphic site. The including step (c) optionally comprises administering the drug or the treatment to the individual if the nucleic acid sample contains the polymorphic variation associated with a positive response to the treatment or the drug and the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug.

[0107] Also provided herein is a method of partnering between a diagnostic/prognostic testing provider and a provider of a consumable product, which comprises: (a) the diagnostic/prognostic testing provider detects the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) the diagnostic/prognostic testing provider identifies the subpopulation of subjects in which the polymorphic variation is associated with osteoarthritis; (c) the diagnostic/prognostic testing provider forwards information to the subpopulation of subjects about a particular product which may be obtained and consumed or applied by the subject to help prevent or delay onset of the disease or condition; and (d) the provider of a consumable product forwards to the diagnostic test provider a fee every time the diagnostic/prognostic test provider forwards information to the subject as set forth in step (c) above.

Compositions Comprising Osteoarthritis-Directed Molecules

[0108] Featured herein is a composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and one or more molecules specifically directed and targeted to a nucleic acid

comprising a *ADAMTS2* nucleotide sequence or amino acid sequence. Such directed molecules include, but are not limited to, a compound that binds to a *ADAMTS2* nucleotide sequence or amino acid sequence referenced herein; a RNAi or siRNA molecule having a strand complementary or substantially complementary to a *ADAMTS2* nucleotide sequence (e.g., hybridizes to a *ADAMTS2* nucleotide sequence under conditions of high stringency); an antisense nucleic acid complementary or substantially complementary to an RNA encoded by a *ADAMTS2* nucleotide sequence (e.g., hybridizes to a *ADAMTS2* nucleotide sequence under conditions of high stringency); a ribozyme that hybridizes to a *ADAMTS2* nucleotide sequence (e.g., hybridizes to a *ADAMTS2* nucleotide sequence under conditions of high stringency); a nucleic acid aptamer that specifically binds a polypeptide encoded by *ADAMTS2* nucleotide sequence; and an antibody that specifically binds to a polypeptide encoded by *ADAMTS2* nucleotide sequence or binds to a nucleic acid having such a nucleotide sequence. In specific embodiments, the osteoarthritis directed molecule interacts with a nucleic acid or polypeptide variant associated with osteoarthritis, such as variants referenced herein. In other embodiments, the osteoarthritis directed molecule interacts with a polypeptide involved in a signal pathway of a polypeptide encoded by a *ADAMTS2* nucleotide sequence, or a nucleic acid comprising such a nucleotide sequence.

[0109] Compositions sometimes include an adjuvant known to stimulate an immune response, and in certain embodiments, an adjuvant that stimulates a T-cell lymphocyte response. Adjuvants are known, including but not limited to an aluminum adjuvant (e.g., aluminum hydroxide); a cytokine adjuvant or adjuvant that stimulates a cytokine response (e.g., interleukin (IL)-12 and/or gamma-interferon cytokines); a Freund-type mineral oil adjuvant emulsion (e.g., Freund's complete or incomplete adjuvant); a synthetic lipid compound; a copolymer adjuvant (e.g., TitreMax); a saponin; Quil A; a liposome; an oil-in-water emulsion (e.g., an emulsion stabilized by Tween 80 and pluronic polyoxyethylene/polyoxypropylene block copolymer (Syntex Adjuvant Formulation); TitreMax; detoxified endotoxin (MPL) and mycobacterial cell wall components (TDW, CWS) in 2% squalene (Ribi Adjuvant System)); a muramyl dipeptide; an immune-stimulating complex (ISCOM, e.g., an Ag-modified saponin/cholesterol micelle that forms stable cage-like structure); an aqueous phase adjuvant that does not have a depot effect (e.g., Gerbu adjuvant); a carbohydrate polymer (e.g., AdjuPrime); L-tyrosine; a manide-oleate compound (e.g., Montanide); an ethylene-vinyl acetate copolymer (e.g., Elvax 40W1,2); or lipid A, for example. Such compositions are useful for generating an immune response against osteoarthritis directed molecule (e.g., an HLA-binding subsequence within a polypeptide encoded by a *ADAMTS2* nucleotide sequence). In such methods, a peptide having an amino acid subsequence of a polypeptide encoded by a *ADAMTS2* nucleotide sequence is delivered to a subject, where the subsequence binds to an HLA molecule and induces a CTL lymphocyte response. The peptide sometimes is delivered to the subject as an isolated peptide or as a minigene in a plasmid that encodes the

peptide. Methods for identifying HLA-binding subsequences in such polypeptides are known (see e.g., publication WO02/20616 and PCT application US98/01373 for methods of identifying such sequences).

[0110] The cell may be in a group of cells cultured *in vitro* or in a tissue maintained *in vitro* or present in an animal *in vivo* (e.g., a rat, mouse, ape or human). In certain embodiments, a composition comprises a component from a cell such as a nucleic acid molecule (e.g., genomic DNA), a protein mixture or isolated protein, for example. The aforementioned compositions have utility in diagnostic, prognostic and pharmacogenomic methods described previously and in therapeutics described hereafter. Certain osteoarthritis directed molecules are described in greater detail below.

Compounds

[0111] Compounds can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive (see, e.g., Zuckermann et al., J. Med. Chem. 37: 2678-85 (1994)); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; "one-bead one-compound" library methods; and synthetic library methods using affinity chromatography selection. Biological library and peptoid library approaches are typically limited to peptide libraries, while the other approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, Anticancer Drug Des. 12: 145, (1997)). Examples of methods for synthesizing molecular libraries are described, for example, in DeWitt et al., Proc. Natl. Acad. Sci. U.S.A. 90: 6909 (1993); Erb et al., Proc. Natl. Acad. Sci. USA 91: 11422 (1994); Zuckermann et al., J. Med. Chem. 37: 2678 (1994); Cho et al., Science 261: 1303 (1993); Carrell et al., Angew. Chem. Int. Ed. Engl. 33: 2059 (1994); Carrell et al., Angew. Chem. Int. Ed. Engl. 33: 2061 (1994); and in Gallop et al., J. Med. Chem. 37: 1233 (1994).

[0112] Libraries of compounds may be presented in solution (e.g., Houghten, Biotechniques 13: 412-421 (1992)), or on beads (Lam, Nature 354: 82-84 (1991)), chips (Fodor, Nature 364: 555-556 (1993)), bacteria or spores (Ladner, United States Patent No. 5,223,409), plasmids (Cull et al., Proc. Natl. Acad. Sci. USA 89: 1865-1869 (1992)) or on phage (Scott and Smith, Science 249: 386-390 (1990); Devlin, Science 249: 404-406 (1990); Cwirla et al., Proc. Natl. Acad. Sci. 87: 6378-6382 (1990); Felici, J. Mol. Biol. 222: 301-310 (1991); Ladner supra.).

[0113] A compound sometimes alters expression and sometimes alters activity of a polypeptide target and may be a small molecule. Small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole,

organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

Antisense Nucleic Acid Molecules, Ribozymes, RNAi, siRNA and Modified Nucleic Acid Molecules

[0114] An "antisense" nucleic acid refers to a nucleotide sequence complementary to a "sense" nucleic acid encoding a polypeptide, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire coding strand, or to a portion thereof or a substantially identical sequence thereof. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence (e.g., 5' and 3' untranslated regions in SEQ ID NO: 1).

[0115] An antisense nucleic acid can be designed such that it is complementary to the entire coding region of an mRNA encoded by a nucleotide sequence (e.g., SEQ ID NO: 1), and often the antisense nucleic acid is an oligonucleotide antisense to only a portion of a coding or noncoding region of the mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of the mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length. The antisense nucleic acids, which include the ribozymes described hereafter, can be designed to target a *ADAMTS2* nucleotide sequence, often a variant associated with osteoarthritis, or a substantially identical sequence thereof. Among the variants, minor alleles and major alleles can be targeted, and those associated with a higher risk of osteoarthritis are often designed, tested, and administered to subjects.

[0116] An antisense nucleic acid can be constructed using chemical synthesis and enzymatic ligation reactions using standard procedures. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[0117] When utilized as therapeutics, antisense nucleic acids typically are administered to a subject (e.g., by direct injection at a tissue site) or generated in situ such that they hybridize with or bind to

cellular mRNA and/or genomic DNA encoding a polypeptide and thereby inhibit expression of the polypeptide, for example, by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then are administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, for example, by linking antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. Antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. Sufficient intracellular concentrations of antisense molecules are achieved by incorporating a strong promoter, such as a pol II or pol III promoter, in the vector construct.

[0118] Antisense nucleic acid molecules sometimes are alpha-anomeric nucleic acid molecules. An alpha-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual beta-units, the strands run parallel to each other (Gaultier et al., *Nucleic Acids. Res.* 15: 6625-6641 (1987)). Antisense nucleic acid molecules can also comprise a 2'-O-methylribonucleotide (Inoue et al., *Nucleic Acids Res.* 15: 6131-6148 (1987)) or a chimeric RNA-DNA analogue (Inoue et al., *FEBS Lett.* 215: 327-330 (1987)). Antisense nucleic acids sometimes are composed of DNA or PNA or any other nucleic acid derivatives described previously.

[0119] In another embodiment, an antisense nucleic acid is a ribozyme. A ribozyme having specificity for a *ADAMT2* nucleotide sequence can include one or more sequences complementary to such a nucleotide sequence, and a sequence having a known catalytic region responsible for mRNA cleavage (see e.g., U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach, *Nature* 334: 585-591 (1988)). For example, a derivative of a Tetrahymena L-19 IVS RNA is sometimes utilized in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a mRNA (see e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742). Also, target mRNA sequences can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (see e.g., Bartel & Szostak, *Science* 261: 1411-1418 (1993)).

[0120] Osteoarthritis directed molecules include in certain embodiments nucleic acids that can form triple helix structures with a *ADAMT2* nucleotide sequence, or a substantially identical sequence thereof, especially one that includes a regulatory region that controls expression of a polypeptide. Gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of a nucleotide sequence referenced herein or a substantially identical sequence (e.g., promoter and/or enhancers) to form triple helical structures that prevent transcription of a gene in target cells (see e.g., Helene, *Anticancer Drug Des.* 6(6): 569-84 (1991); Helene et al., *Ann. N.Y. Acad. Sci.* 660: 27-36 (1992); and Maher, *Bioassays* 14(12): 807-15 (1992). Potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one

strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

[0121] Osteoarthritis directed molecules include RNAi and siRNA nucleic acids. Gene expression may be inhibited by the introduction of double-stranded RNA (dsRNA), which induces potent and specific gene silencing, a phenomenon called RNA interference or RNAi. See, e.g., Fire et al., US Patent Number 6,506,559; Tuschl et al. PCT International Publication No. WO 01/75164; Kay et al. PCT International Publication No. WO 03/010180A1; or Bosher JM, Labouesse, Nat Cell Biol 2000 Feb;2(2):E31-6. This process has been improved by decreasing the size of the double-stranded RNA to 20-24 base pairs (to create small-interfering RNAs or siRNAs) that “switched off” genes in mammalian cells without initiating an acute phase response, i.e., a host defense mechanism that often results in cell death (see, e.g., Caplen et al. Proc Natl Acad Sci U S A. 2001 Aug 14;98(17):9742-7 and Elbashir et al. Methods 2002 Feb;26(2):199-213). There is increasing evidence of post-transcriptional gene silencing by RNA interference (RNAi) for inhibiting targeted expression in mammalian cells at the mRNA level, in human cells. There is additional evidence of effective methods for inhibiting the proliferation and migration of tumor cells in human patients, and for inhibiting metastatic cancer development (see, e.g., U.S. Patent Application No. US2001000993183; Caplen et al. Proc Natl Acad Sci U S A; and Abderrahmani et al. Mol Cell Biol 2001 Nov21(21):7256-67).

[0122] An “siRNA” or “RNAi” refers to a nucleic acid that forms a double stranded RNA and has the ability to reduce or inhibit expression of a gene or target gene when the siRNA is delivered to or expressed in the same cell as the gene or target gene. “siRNA” refers to short double-stranded RNA formed by the complementary strands. Complementary portions of the siRNA that hybridize to form the double stranded molecule often have substantial or complete identity to the target molecule sequence. In one embodiment, an siRNA refers to a nucleic acid that has substantial or complete identity to a target gene and forms a double stranded siRNA.

[0123] When designing the siRNA molecules, the targeted region often is selected from a given DNA sequence beginning 50 to 100 nucleotides downstream of the start codon. See, e.g., Elbashir et al., Methods 26:199-213 (2002). Initially, 5' or 3' UTRs and regions nearby the start codon were avoided assuming that UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP or RISC endonuclease complex. Sometimes regions of the target 23 nucleotides in length conforming to the sequence motif AA(N19)TT (N, an nucleotide), and regions with approximately 30% to 70% G/C-content (often about 50% G/C-content) often are selected. If no suitable sequences are found, the search often is extended using the motif NA(N21). The sequence of the sense siRNA sometimes corresponds to (N19) TT or N21 (position 3 to 23 of the 23-nt motif), respectively. In the latter case, the 3' end of the sense siRNA often is converted to TT. The rationale for this sequence conversion is to generate a symmetric duplex with respect to the sequence composition of the sense and

antisense 3' overhangs. The antisense siRNA is synthesized as the complement to position 1 to 21 of the 23-nt motif. Because position 1 of the 23-nt motif is not recognized sequence-specifically by the antisense siRNA, the 3'-most nucleotide residue of the antisense siRNA can be chosen deliberately. However, the penultimate nucleotide of the antisense siRNA (complementary to position 2 of the 23-nt motif) often is complementary to the targeted sequence. For simplifying chemical synthesis, TT often is utilized. siRNAs corresponding to the target motif NAR(N17)YNN, where R is purine (A,G) and Y is pyrimidine (C,U), often are selected. Respective 21 nucleotide sense and antisense siRNAs often begin with a purine nucleotide and can also be expressed from pol III expression vectors without a change in targeting site. Expression of RNAs from pol III promoters often is efficient when the first transcribed nucleotide is a purine.

[0124] The sequence of the siRNA can correspond to the full length target gene, or a subsequence thereof. Often, the siRNA is about 15 to about 50 nucleotides in length (e.g., each complementary sequence of the double stranded siRNA is 15-50 nucleotides in length, and the double stranded siRNA is about 15-50 base pairs in length, sometimes about 20-30 nucleotides in length or about 20-25 nucleotides in length, e.g., 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length. The siRNA sometimes is about 21 nucleotides in length. Methods of using siRNA are well known in the art, and specific siRNA molecules may be purchased from a number of companies including Dharmacon Research, Inc.

[0125] Antisense, ribozyme, RNAi and siRNA nucleic acids can be altered to form modified nucleic acid molecules. The nucleic acids can be altered at base moieties, sugar moieties or phosphate backbone moieties to improve stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup et al., *Bioorganic & Medicinal Chemistry* 4 (1): 5-23 (1996)). As used herein, the terms "peptide nucleic acid" or "PNA" refers to a nucleic acid mimic such as a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. Synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described, for example, in Hyrup et al., (1996) *supra* and Perry-O'Keefe et al., *Proc. Natl. Acad. Sci.* 93: 14670-675 (1996).

[0126] PNA nucleic acids can be used in prognostic, diagnostic, and therapeutic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNA nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as "artificial restriction enzymes" when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup et al., (1996) *supra*; Perry-O'Keefe *supra*).

[0127] In other embodiments, oligonucleotides may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across cell membranes (see e.g., Letsinger et al., Proc. Natl. Acad. Sci. USA 86: 6553-6556 (1989); Lemaitre et al., Proc. Natl. Acad. Sci. USA 84: 648-652 (1987); PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, e.g., Krol et al., Bio-Techniques 6: 958-976 (1988)) or intercalating agents. (See, e.g., Zon, Pharm. Res. 5: 539-549 (1988)). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[0128] Also included herein are molecular beacon oligonucleotide primer and probe molecules having one or more regions complementary to a *ADAMTS2* nucleotide sequence, or a substantially identical sequence thereof, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantifying the presence of the nucleic acid in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi et al., U.S. Patent No. 5,854,033; Nazarenko et al., U.S. Patent No. 5,866,336, and Livak et al., U.S. Patent 5,876,930.

Antibodies

[0129] The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')₂; fragments which can be generated by treating the antibody with an enzyme such as pepsin. An antibody sometimes is a polyclonal, monoclonal, recombinant (e.g., a chimeric or humanized), fully human, non-human (e.g., murine), or a single chain antibody. An antibody may have effector function and can fix complement, and is sometimes coupled to a toxin or imaging agent.

[0130] A full-length polypeptide or antigenic peptide fragment encoded by a nucleotide sequence referenced herein can be used as an immunogen or can be used to identify antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. An antigenic peptide often includes at least 8 amino acid residues of the amino acid sequences encoded by a nucleotide sequence referenced herein, or substantially identical sequence thereof, and encompasses an epitope. Antigenic peptides sometimes include 10 or more amino acids, 15 or more amino acids, 20 or more amino acids, or 30 or more amino acids. Hydrophilic and hydrophobic fragments of polypeptides sometimes are used as immunogens.

[0131] Epitopes encompassed by the antigenic peptide are regions located on the surface of the polypeptide (e.g., hydrophilic regions) as well as regions with high antigenicity. For example, an Ermini surface probability analysis of the human polypeptide sequence can be used to indicate the regions that

have a particularly high probability of being localized to the surface of the polypeptide and are thus likely to constitute surface residues useful for targeting antibody production. The antibody may bind an epitope on any domain or region on polypeptides described herein.

[0132] Also, chimeric, humanized, and completely human antibodies are useful for applications which include repeated administration to subjects. Chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, can be made using standard recombinant DNA techniques. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson et al International Application No. PCT/US86/02269; Akira, et al European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al European Patent Application 173,494; Neuberger et al PCT International Publication No. WO 86/01533; Cabilly et al U.S. Patent No. 4,816,567; Cabilly et al European Patent Application 125,023; Better et al., Science 240: 1041-1043 (1988); Liu et al., Proc. Natl. Acad. Sci. USA 84: 3439-3443 (1987); Liu et al., J. Immunol. 139: 3521-3526 (1987); Sun et al., Proc. Natl. Acad. Sci. USA 84: 214-218 (1987); Nishimura et al., Canc. Res. 47: 999-1005 (1987); Wood et al., Nature 314: 446-449 (1985); and Shaw et al., J. Natl. Cancer Inst. 80: 1553-1559 (1988); Morrison, S. L., Science 229: 1202-1207 (1985); Oi et al., BioTechniques 4: 214 (1986); Winter U.S. Patent 5,225,539; Jones et al., Nature 321: 552-525 (1986); Verhoeven et al., Science 239: 1534; and Beidler et al., J. Immunol. 141: 4053-4060 (1988).

[0133] Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice that are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. See, for example, Lonberg and Huszar, Int. Rev. Immunol. 13: 65-93 (1995); and U.S. Patent Nos. 5,625,126; 5,633,425; 5,569,825; 5,661,016; and 5,545,806. In addition, companies such as Abgenix, Inc. (Fremont, CA) and Medarex, Inc. (Princeton, NJ), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above. Completely human antibodies that recognize a selected epitope also can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody (e.g., a murine antibody) is used to guide the selection of a completely human antibody recognizing the same epitope. This technology is described for example by Jespers et al., Bio/Technology 12: 899-903 (1994).

[0134] An antibody can be a single chain antibody. A single chain antibody (scFv) can be engineered (see, e.g., Colcher et al., Ann. N Y Acad. Sci. 880: 263-80 (1999); and Reiter, Clin. Cancer Res. 2: 245-52 (1996)). Single chain antibodies can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target polypeptide.

[0135] Antibodies also may be selected or modified so that they exhibit reduced or no ability to bind an Fc receptor. For example, an antibody may be an isotype or subtype, fragment or other mutant, which

does not support binding to an Fc receptor (e.g., it has a mutagenized or deleted Fc receptor binding region).

[0136] Also, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1 dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thiotepa chlorambucil, melphalan, carmustine (BCNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine and vinblastine).

[0137] Antibody conjugates can be used for modifying a given biological response. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a polypeptide such as tumor necrosis factor, gamma-interferon, alpha-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors. Also, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980, for example.

[0138] An antibody (e.g., monoclonal antibody) can be used to isolate target polypeptides by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an antibody can be used to detect a target polypeptide (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. Antibodies can be used diagnostically to monitor polypeptide levels in tissue as part of a clinical testing procedure, e.g., to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labeling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials

include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H . Also, an antibody can be utilized as a test molecule for determining whether it can treat osteoarthritis, and as a therapeutic for administration to a subject for treating osteoarthritis.

[0139] An antibody can be made by immunizing with a purified antigen, or a fragment thereof, e.g., a fragment described herein, a membrane associated antigen, tissues, e.g., crude tissue preparations, whole cells, preferably living cells, lysed cells, or cell fractions.

[0140] Included herein are antibodies which bind only a native polypeptide, only denatured or otherwise non-native polypeptide, or which bind both, as well as those having linear or conformational epitopes. Conformational epitopes sometimes can be identified by selecting antibodies that bind to native but not denatured polypeptide. Also featured are antibodies that specifically bind to a polypeptide variant associated with osteoarthritis.

Methods for Identifying Candidate Therapeutics for Treating Osteoarthritis

[0141] Current therapies for the treatment of osteoarthritis have limited efficacy, limited tolerability and significant mechanism-based side effects, and few of the available therapies adequately address underlying defects. Current therapeutic approaches were largely developed in the absence of defined molecular targets or even a solid understanding of disease pathogenesis. Therefore, provided are methods of identifying candidate therapeutics that target biochemical pathways related to the development of osteoarthritis.

[0142] Thus, featured herein are methods for identifying a candidate therapeutic for treating osteoarthritis. The methods comprise contacting a test molecule with a target molecule in a system. A "target molecule" as used herein refers to a *ADAMTS2* nucleic acid, a substantially identical nucleic acid thereof, or a fragment thereof, and an encoded polypeptide of the foregoing. The methods also comprise determining the presence or absence of an interaction between the test molecule and the target molecule, where the presence of an interaction between the test molecule and the nucleic acid or polypeptide identifies the test molecule as a candidate osteoarthritis therapeutic. The interaction between the test molecule and the target molecule may be quantified.

[0143] Test molecules and candidate therapeutics include, but are not limited to, compounds, antisense nucleic acids, siRNA molecules, ribozymes, polypeptides or proteins encoded by a *ADAMTS2* nucleotide sequence, or a substantially identical sequence or fragment thereof, and immunotherapeutics (e.g., antibodies and HLA-presented polypeptide fragments). A test molecule or candidate therapeutic may act as a modulator of target molecule concentration or target molecule function in a system. A

“modulator” may agonize (i.e., up-regulates) or antagonize (i.e., down-regulates) a target molecule concentration partially or completely in a system by affecting such cellular functions as DNA replication and/or DNA processing (e.g., DNA methylation or DNA repair), RNA transcription and/or RNA processing (e.g., removal of intronic sequences and/or translocation of spliced mRNA from the nucleus), polypeptide production (e.g., translation of the polypeptide from mRNA), and/or polypeptide post-translational modification (e.g., glycosylation, phosphorylation, and proteolysis of pro-polypeptides). A modulator may also agonize or antagonize a biological function of a target molecule partially or completely, where the function may include adopting a certain structural conformation, interacting with one or more binding partners, ligand binding, catalysis (e.g., phosphorylation, dephosphorylation, hydrolysis, methylation, and isomerization), and an effect upon a cellular event (e.g., effecting progression of osteoarthritis).

[0144] As used herein, the term “system” refers to a cell free *in vitro* environment and a cell-based environment such as a collection of cells, a tissue, an organ, or an organism. A system is “contacted” with a test molecule in a variety of manners, including adding molecules in solution and allowing them to interact with one another by diffusion, cell injection, and any administration routes in an animal. As used herein, the term “interaction” refers to an effect of a test molecule on test molecule, where the effect sometimes is binding between the test molecule and the target molecule, and sometimes is an observable change in cells, tissue, or organism.

[0145] There are many standard methods for detecting the presence or absence of interaction between a test molecule and a target molecule. For example, titrametric, acidimetric, radiometric, NMR, monolayer, polarographic, spectrophotometric, fluorescent, and ESR assays probative of a target molecule interaction may be utilized.

[0146] *ADAMTS2* activity and/or *ADAMTS2* interactions can be detected and quantified using assays known in the art. For example, an immunoprecipitation assay or a kinase activity assay that employs a kinase-inactivated MEK can be utilized. Kinase inactivated MEKs are known in the art, such as a MEK that includes the mutation K97M. In these assays, mammalian cells (e.g., COS or NIH-3T3) are transiently transfected with constructs expressing *ADAMTS2*, and in addition, the cells are co-transfected with oncogenic RAS or SRC or both. Oncogenic RAS or SRC activates *ADAMTS2* kinase activity. *ADAMTS2* is immunoprecipitated from cell extracts using a monoclonal antibody (e.g., 9E10) or a polyclonal antibody (e.g., from rabbit) specific for a unique peptide from *ADAMTS2*. *ADAMTS2* is then resuspended in assay buffer containing GST-Mek1 or GST-Mek2 and/or GST-ERK2. In addition, [γ 32 P] ATP can be added to detect and/or quantify phosphorylation activity. Samples are incubated for 5-30 minutes at 30°C, and then the reaction is terminated by addition of EDTA. The samples are centrifuged and the supernatant fractions are collected. Phosphorylation activity is detected using one of two methods: (i) activity of GST-ERK2 kinase can be measured using MBP (myelin basic

protein, a substrate for ERK) as substrate, or (ii) following incubation of immunoprecipitated *ADAMTS2* in reaction buffer containing GST-ERK and [γ - ^{32}P] ATP, transfer of labeled ATP to kinase-dead ERK can be quantified by a phosphor-imager or densitometer following PAGE separation of polypeptide products (phosphorylated and non-phosphorylated forms). These types of assays are described in Weber et al., *Oncogene* 19: 169-176 (2000); Mason et al., *EMBO J.* 18: 2137-2148 (1999); Marais et al., *J. Biol. Chem.* 272: 4378-4383 (1997); Marais et al., *EMBO J.* 14: 3136-3145 (1995).

[0147] As noted above, *ADAMTS2* includes a domain having metalloprotease activity, and modulators of such activity are known. Examples of such modulators are set forth in WO03063762A2; WO-09937625; WO-09918076; WO-09838163; WO-09837877; WO9947550A1; WO0177092A1; WO0040577A1; WO9942436A1; WO9838163A1; WO9837877A1; WO04014379A1; WO03106381A2; WO03014098A1; WO03014092A1 and WO02096426A1.

[0148] Test molecule/target molecule interactions can be detected and/or quantified using assays known in the art. For example, an interaction can be determined by labeling the test molecule and/or the target molecule, where the label is covalently or non-covalently attached to the test molecule or target molecule. The label is sometimes a radioactive molecule such as ^{125}I , ^{131}I , ^{35}S or ^3H , which can be detected by direct counting of radioemission or by scintillation counting. Also, enzymatic labels such as horseradish peroxidase, alkaline phosphatase, or luciferase may be utilized where the enzymatic label can be detected by determining conversion of an appropriate substrate to product. In addition, presence or absence of an interaction can be determined without labeling. For example, a microphysiometer (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indication of an interaction between a test molecule and target molecule (McConnell, H. M. *et al.*, *Science* 257: 1906-1912 (1992)).

[0149] In cell-based systems, cells typically include a *ADAMTS2* nucleic acid, an encoded polypeptide, or substantially identical nucleic acid or polypeptide thereof, and are often of mammalian origin, although the cell can be of any origin. Whole cells, cell homogenates, and cell fractions (e.g., cell membrane fractions) can be subjected to analysis. Where interactions between a test molecule with a target polypeptide are monitored, soluble and/or membrane bound forms of the polypeptide may be utilized. Where membrane-bound forms of the polypeptide are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton[®] X-100, Triton[®] X-114, Thesit[®], Isotridecypoly(ethylene glycol ether)_n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-

cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

[0150] An interaction between a test molecule and target molecule also can be detected by monitoring fluorescence energy transfer (FET) (*see, e.g.,* Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos *et al.* U.S. Patent No. 4,868,103). A fluorophore label on a first, “donor” molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, “acceptor” molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the “donor” polypeptide molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the “acceptor” molecule label may be differentiated from that of the “donor”. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the “acceptor” molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (*e.g.,* using a fluorimeter).

[0151] In another embodiment, determining the presence or absence of an interaction between a test molecule and a target molecule can be effected by monitoring surface plasmon resonance (*see, e.g.,* Sjolander & Urbanicz, *Anal. Chem.* 63: 2338-2345 (1991) and Szabo *et al., Curr. Opin. Struct. Biol.* 5: 699-705 (1995)). “Surface plasmon resonance” or “biomolecular interaction analysis (BIA)” can be utilized to detect biospecific interactions in real time, without labeling any of the interactants (*e.g.,* BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

[0152] In another embodiment, the target molecule or test molecules are anchored to a solid phase, facilitating the detection of target molecule/test molecule complexes and separation of the complexes from free, uncomplexed molecules. The target molecule or test molecule is immobilized to the solid support. In an embodiment, the target molecule is anchored to a solid surface, and the test molecule, which is not anchored, can be labeled, either directly or indirectly, with detectable labels discussed herein.

[0153] It may be desirable to immobilize a target molecule, an anti-target molecule antibody, and/or test molecules to facilitate separation of target molecule/test molecule complexes from uncomplexed forms, as well as to accommodate automation of the assay. The attachment between a test molecule and/or target molecule and the solid support may be covalent or non-covalent (*see, e.g.,* U.S. Patent No. 6,022,688 for non-covalent attachments). The solid support may be one or more surfaces of the system, such as one or more surfaces in each well of a microtiter plate, a surface of a silicon wafer, a surface of a

bead (*see, e.g., Lam, Nature 354: 82-84 (1991)*) that is optionally linked to another solid support, or a channel in a microfluidic device, for example. Types of solid supports, linker molecules for covalent and non-covalent attachments to solid supports, and methods for immobilizing nucleic acids and other molecules to solid supports are well known (*see, e.g., U.S. Patent Nos. 6,261,776; 5,900,481; 6,133,436; and 6,022,688; and WIPO publication WO 01/18234*).

[0154] In an embodiment, target molecule may be immobilized to surfaces via biotin and streptavidin. For example, biotinylated target polypeptide can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g., biotinylation kit, Pierce Chemicals, Rockford, IL*), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). In another embodiment, a target polypeptide can be prepared as a fusion polypeptide. For example, glutathione-S-transferase/target polypeptide fusion can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivitized microtiter plates, which are then combined with a test molecule under conditions conducive to complex formation (*e.g., at physiological conditions for salt and pH*). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, or the matrix is immobilized in the case of beads, and complex formation is determined directly or indirectly as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of target molecule binding or activity is determined using standard techniques.

[0155] In an embodiment, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (*e.g., by washing*) under conditions such that a significant percentage of complexes formed will remain immobilized to the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of manners. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface, *e.g., by adding a labeled antibody specific for the immobilized component, where the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody*.

[0156] In another embodiment, an assay is performed utilizing antibodies that specifically bind target molecule or test molecule but do not interfere with binding of the target molecule to the test molecule. Such antibodies can be derivitized to a solid support, and unbound target molecule may be immobilized by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

[0157] Cell free assays also can be conducted in a liquid phase. In such an assay, reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (*see, e.g., Rivas, G., and Minton, Trends Biochem Sci Aug;18(8): 284-7 (1993)*); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (*see, e.g., Ausubel et al., eds. Current Protocols in Molecular Biology, J. Wiley: New York (1999)*); and immunoprecipitation (*see, e.g., Ausubel et al., eds., supra*). Media and chromatographic techniques are known to one skilled in the art (*see, e.g., Heegaard, J Mol. Recognit. Winter; 11(1-6): 141-8 (1998)*; Hage & Tweed, *J. Chromatogr. B Biomed. Sci. Appl. Oct 10: 699 (1-2): 499-525 (1997)*). Further, fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

[0158] In another embodiment, modulators of target molecule expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of target mRNA or target polypeptide is evaluated relative to the level of expression of target mRNA or target polypeptide in the absence of the candidate compound. When expression of target mRNA or target polypeptide is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as an agonist of target mRNA or target polypeptide expression. Alternatively, when expression of target mRNA or target polypeptide is less (*e.g., less with statistical significance*) in the presence of the candidate compound than in its absence, the candidate compound is identified as an antagonist or inhibitor of target mRNA or target polypeptide expression. The level of target mRNA or target polypeptide expression can be determined by methods described herein.

[0159] In another embodiment, binding partners that interact with a target molecule are detected. The target molecules can interact with one or more cellular or extracellular macromolecules, such as polypeptides *in vivo*, and these interacting molecules are referred to herein as "binding partners." Binding partners can agonize or antagonize target molecule biological activity. Also, test molecules that agonize or antagonize interactions between target molecules and binding partners can be useful as therapeutic molecules as they can up-regulate or down-regulated target molecule activity *in vivo* and thereby treat osteoarthritis.

[0160] Binding partners of target molecules can be identified by methods known in the art. For example, binding partners may be identified by lysing cells and analyzing cell lysates by electrophoretic techniques. Alternatively, a two-hybrid assay or three-hybrid assay can be utilized (*see, e.g., U.S. Patent No. 5,283,317; Zervos et al., Cell 72:223-232 (1993)*; Madura *et al., J. Biol. Chem.* 268: 12046-12054 (1993); Bartel *et al., Biotechniques* 14: 920-924 (1993); Iwabuchi *et al., Oncogene* 8: 1693-1696 (1993); and Brent WO94/10300). A two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. The assay often utilizes two different DNA constructs. In one construct, a *ADAMTS2* nucleic acid (sometimes referred to as the

“bait”) is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In another construct, a DNA sequence from a library of DNA sequences that encodes a potential binding partner (sometimes referred to as the “prey”) is fused to a gene that encodes an activation domain of the known transcription factor. Sometimes, a *ADAMTS2* nucleic acid can be fused to the activation domain. If the “bait” and the “prey” molecules interact *in vivo*, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to identify the potential binding partner.

[0161] In an embodiment for identifying test molecules that antagonize or agonize complex formation between target molecules and binding partners, a reaction mixture containing the target molecule and the binding partner is prepared, under conditions and for a time sufficient to allow complex formation. The reaction mixture often is provided in the presence or absence of the test molecule. The test molecule can be included initially in the reaction mixture, or can be added at a time subsequent to the addition of the target molecule and its binding partner. Control reaction mixtures are incubated without the test molecule or with a placebo. Formation of any complexes between the target molecule and the binding partner then is detected. Decreased formation of a complex in the reaction mixture containing test molecule as compared to in a control reaction mixture indicates that the molecule antagonizes target molecule/binding partner complex formation. Alternatively, increased formation of a complex in the reaction mixture containing test molecule as compared to in a control reaction mixture indicates that the molecule agonizes target molecule/binding partner complex formation. In another embodiment, complex formation of target molecule/binding partner can be compared to complex formation of mutant target molecule/binding partner (*e.g.*, amino acid modifications in a target polypeptide). Such a comparison can be important in those cases where it is desirable to identify test molecules that modulate interactions of mutant but not non-mutated target gene products.

[0162] The assays can be conducted in a heterogeneous or homogeneous format. In heterogeneous assays, target molecule and/or the binding partner are immobilized to a solid phase, and complexes are detected on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the molecules being tested. For example, test compounds that agonize target molecule/binding partner interactions can be identified by conducting the reaction in the presence of the test molecule in a competition format. Alternatively, test molecules that agonize preformed complexes, *e.g.*, molecules with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed.

[0163] In a heterogeneous assay embodiment, the target molecule or the binding partner is anchored onto a solid surface (e.g., a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored molecule can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the molecule to be anchored can be used to anchor the molecule to the solid surface. The partner of the immobilized species is exposed to the coated surface with or without the test molecule. After the reaction is complete, unreacted components are removed (e.g., by washing) such that a significant portion of any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface is indicative of complex. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored to the surface; e.g., by using a labeled antibody specific for the initially non-immobilized species. Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

[0164] In another embodiment, the reaction can be conducted in a liquid phase in the presence or absence of test molecule, where the reaction products are separated from unreacted components, and the complexes are detected (e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes). Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

[0165] In an alternate embodiment, a homogeneous assay can be utilized. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared. One or both of the target molecule or binding partner is labeled, and the signal generated by the label(s) is quenched upon complex formation (e.g., U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). Addition of a test molecule that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target molecule/binding partner complexes can be identified.

[0166] Candidate therapeutics for treating osteoarthritis are identified from a group of test molecules that interact with a target molecule. Test molecules are normally ranked according to the degree with which they modulate (e.g., agonize or antagonize) a function associated with the target molecule (e.g., DNA replication and/or processing, RNA transcription and/or processing, polypeptide production and/or processing, and/or biological function/activity), and then top ranking modulators are selected. Also, pharmacogenomic information described herein can determine the rank of a modulator. The top 10% of ranked test molecules often are selected for further testing as candidate therapeutics, and sometimes the top 15%, 20%, or 25% of ranked test molecules are selected for further testing as candidate therapeutics. Candidate therapeutics typically are formulated for administration to a subject.

Therapeutic Formulations

[0167] Formulations and pharmaceutical compositions typically include in combination with a pharmaceutically acceptable carrier one or more target molecule modulators. The modulator often is a test molecule identified as having an interaction with a target molecule by a screening method described above. The modulator may be a compound, an antisense nucleic acid, a ribozyme, an antibody, or a binding partner. Also, formulations may comprise a target polypeptide or fragment thereof in combination with a pharmaceutically acceptable carrier.

[0168] Formulations or pharmaceutical compositions typically include in combination with a pharmaceutically acceptable carrier, a compound, an antisense nucleic acid, a ribozyme, an antibody, a binding partner that interacts with an *ADAMTS2* polypeptide, a *ADAMTS2* nucleic acid, or a fragment thereof. The formulated molecule may be one that is identified by a screening method described above. Also, formulations may comprise a *ADAMTS2* polypeptide or fragment thereof, where the *ADAMTS2* polypeptide contains an isoleucine at position 245 of SEQ ID NO: 4, and a pharmaceutically acceptable carrier. Also, formulations may comprise an active *ADAMTS2* polypeptide or fragment thereof, where *ADAMTS2* polypeptide fragments having activity are selected from amino acids 252-1211, 253-1211, 254-1211, 255-1211, 256-1211, 257-1211, 258-1211, 259-1211 or 260-1211 of SEQ ID NO: 4, where it is understood that the active form of *ADAMTS2* does not contain the propeptide domain. As used herein, the term "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

[0169] As used herein, the term "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions. Pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

[0170] A pharmaceutical composition typically is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerin, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with

acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[0171] Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, e.g., gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[0172] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[0173] Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which

yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0174] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

[0175] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art. Molecules can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[0176] In one embodiment, active molecules are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. Materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[0177] It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

[0178] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Molecules which exhibit high therapeutic indices are preferred. While molecules that exhibit toxic side effects may be used, care should be taken to design a delivery system

that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[0179] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such molecules lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any molecules used in the methods described herein, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[0180] As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, sometimes about 0.01 to 25 mg/kg body weight, often about 0.1 to 20 mg/kg body weight, and more often about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, sometimes between 2 to 8 weeks, often between about 3 to 7 weeks, and more often for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors may influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

[0181] With regard to polypeptide formulations, featured herein is a method for treating osteoarthritis in a subject, which comprises contacting one or more cells in the subject with a first polypeptide, where the subject comprises a second polypeptide having one or more polymorphic variations associated with cancer, and where the first polypeptide comprises fewer polymorphic variations associated with cancer than the second polypeptide. The first and second polypeptides are encoded by a nucleic acid which comprises a nucleotide sequence in SEQ ID NO: 1-3; a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence referenced in SEQ ID NO: 1-3; a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-3 and a nucleotide sequence 90% or more identical to a nucleotide sequence in SEQ ID NO: 1-3. The subject often is a human.

[0182] For antibodies, a dosage of 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg) is often utilized. If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is often appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (*e.g.*, into the brain). A method for lipidation of antibodies is described by Cruikshank *et al.*, *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193 (1997).

[0183] Antibody conjugates can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a polypeptide such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors. Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[0184] For compounds, exemplary doses include milligram or microgram amounts of the compound per kilogram of subject or sample weight, for example, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid described herein, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[0185] With regard to nucleic acid formulations, gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent 5,328,470) or by stereotactic injection (*see e.g.*, Chen *et al.*, (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). Pharmaceutical preparations of gene therapy vectors can include a gene therapy vector in an acceptable

diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells (e.g., retroviral vectors) the pharmaceutical preparation can include one or more cells which produce the gene delivery system. Examples of gene delivery vectors are described herein.

Therapeutic Methods

[0186] A therapeutic formulation described above can be administered to a subject in need of a therapeutic for inducing a desired biological response.. Therapeutic formulations can be administered by any of the paths described herein. With regard to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from pharmacogenomic analyses described herein.

[0187] As used herein, the term "treatment" is defined as the application or administration of a therapeutic formulation to a subject, or application or administration of a therapeutic agent to an isolated tissue or cell line from a subject with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect osteoarthritis, symptoms of osteoarthritis or a predisposition towards osteoarthritis. A therapeutic formulation includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides. Administration of a therapeutic formulation can occur prior to the manifestation of symptoms characteristic of osteoarthritis, such that osteoarthritis is prevented or delayed in its progression. The appropriate therapeutic composition can be determined based on screening assays described herein.

[0188] As discussed, successful treatment of osteoarthritis can be brought about by techniques that serve to agonize target molecule expression or function, or alternatively, antagonize target molecule expression or function. These techniques include administration of modulators that include, but are not limited to, small organic or inorganic molecules; antibodies (including, for example, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and Fab, F(ab')₂ and Fab expression library fragments, scFV molecules, and epitope-binding fragments thereof); and peptides, phosphopeptides, or polypeptides.

[0189] Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above. It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode

and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular polypeptide, it can be preferable to co-administer normal target gene polypeptide into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

[0190] Another method by which nucleic acid molecules may be utilized in treating or preventing osteoarthritis is use of aptamer molecules specific for target molecules. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically bind to ligands (*see, e.g., Osborne, et al., Curr. Opin. Chem. Biol.* 1(1): 5-9 (1997); and Patel, D. J., *Curr. Opin. Chem. Biol. Jun;1(1): 32-46 (1997)*).

[0191] Yet another method of utilizing nucleic acid molecules for osteoarthritis treatment is gene therapy, which can also be referred to as allele therapy. Provided herein is a gene therapy method for treating osteoarthritis in a subject, which comprises contacting one or more cells in the subject or from the subject with a nucleic acid having a first nucleotide sequence (e.g., the first nucleotide sequence is identical to or substantially identical to a nucleotide sequence of SEQ ID NO: 1-3). Genomic DNA in the subject comprises a second nucleotide sequence having one or more polymorphic variations associated with osteoarthritis (e.g., the second nucleotide sequence is identical to or substantially identical to a nucleotide sequence of SEQ ID NO: 1). The first and second nucleotide sequences typically are substantially identical to one another, and the first nucleotide sequence comprises fewer polymorphic variations associated with osteoarthritis than the second nucleotide sequence. The first nucleotide sequence may comprise a gene sequence that encodes a full-length polypeptide or a fragment thereof. The subject is often a human. Allele therapy methods often are utilized in conjunction with a method of first determining whether a subject has genomic DNA that includes polymorphic variants associated with osteoarthritis.

[0192] In another allele therapy embodiment, provided herein is a method which comprises contacting one or more cells in the subject or from the subject with a polypeptide encoded by a nucleic acid having a first nucleotide sequence (e.g., the first nucleotide sequence is identical to or substantially identical to the nucleotide sequence of SEQ ID NO: 1-3). Genomic DNA in the subject comprises a second nucleotide sequence having one or more polymorphic variations associated with osteoarthritis (e.g., the second nucleotide sequence is identical to or substantially identical to a nucleotide sequence of SEQ ID NO: 1). The first and second nucleotide sequences typically are substantially identical to one another, and the first nucleotide sequence comprises fewer polymorphic variations associated with osteoarthritis than the second nucleotide sequence. The first nucleotide sequence may comprise a gene sequence that encodes a full-length polypeptide or a fragment thereof. The subject is often a human. The method often comprises supplementing arthritis-associated *ADAMTS2* polypeptide with a non-arthritis-associated *ADAMTS2* polypeptide or fragment thereof, where the non-arthritis-associated form of

ADAMTS2 contains an isoleucine at position 245 of SEQ ID NO: 4 having enzymatic activity. The arthritis-associated *ADAMTS2* polypeptide sometimes contains a valine at position 245 of SEQ ID NO: 4 having an altered enzymatic activity varying from the non-arthritis-associated polypeptide.

[0193] In an embodiment, provided is a method of increasing the synthesis of procollagen II comprising providing or administering to individuals in need of increasing levels of type II collagen the pharmaceutical or physiologically acceptable composition comprising active human *ADAMTS2* protein or fragment thereof, where *ADAMTS2* polypeptide fragments having activity are selected from amino acids 252-1211, 253-1211, 254-1211, 255-1211, 256-1211, 257-1211, 258-1211, 259-1211 or 260-1211 of SEQ ID NO: 4, where it is understood that the active form of *ADAMTS2* does not contain the propeptide domain.

[0194] In another embodiment, provided herein is a method of increasing the synthesis of procollagen II comprising providing or administering to individuals in need of increasing levels of type II collagen the pharmaceutical or physiologically acceptable composition comprising an enzyme or molecule capable of cleaving *ADAMTS2* propeptide, e.g., a furin-type endopeptidase or N-ethylmaleimide described herein

[0195] For antibody-based therapies, antibodies can be generated that are both specific for target molecules and that reduce target molecule activity. Such antibodies may be administered in instances where antagonizing a target molecule function is appropriate for the treatment of osteoarthritis.

[0196] In circumstances where stimulating antibody production in an animal or a human subject by injection with a target molecule is harmful to the subject, it is possible to generate an immune response against the target molecule by use of anti-idiotypic antibodies (*see, e.g., Herlyn, Ann. Med.;31(1): 66-78 (1999); and Bhattacharya-Chatterjee & Foon, Cancer Treat. Res.; 94: 51-68 (1998)*). Introducing an anti-idiotypic antibody to a mammal or human subject often stimulates production of anti-anti-idiotypic antibodies, which typically are specific to the target molecule. Vaccines directed to osteoarthritis also may be generated in this fashion.

[0197] In instances where the target molecule is intracellular and whole antibodies are used, internalizing antibodies may be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (*see, e.g., Marasco et al., Proc. Natl. Acad. Sci. USA 90: 7889-7893 (1993)*).

[0198] Modulators can be administered to a patient at therapeutically effective doses to treat osteoarthritis. A therapeutically effective dose refers to an amount of the modulator sufficient to result in amelioration of symptoms of osteoarthritis. Toxicity and therapeutic efficacy of modulators can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Modulators that exhibit large therapeutic indices are preferred. While modulators that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such molecules to the site of affected tissue in order to minimize potential damage to uninfected cells, thereby reducing side effects.

[0199] Data obtained from cell culture assays and animal studies can be used in formulating a range of dosages for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the methods described herein, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[0200] Another example of effective dose determination for an individual is the ability to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays may utilize antibody mimics and/or "biosensors" that have been created through molecular imprinting techniques. Molecules that modulate target molecule activity are used as a template, or "imprinting molecule", to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated "negative image" of the compound and is able to selectively rebind the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell *et al.*, *Current Opinion in Biotechnology* 7: 89-94 (1996) and in Shea, *Trends in Polymer Science* 2: 166-173 (1994). Such "imprinted" affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis, *et al.*, *Nature* 361: 645-647 (1993). Through the use of isotope-labeling, the "free" concentration of compound which modulates target molecule expression or activity readily can be monitored and used in calculations of IC₅₀. Such "imprinted" affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties

measurably change upon local and selective binding of target compound. These changes readily can be assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC_{50} . An example of such a "biosensor" is discussed in Kriz *et al.*, *Analytical Chemistry* 67: 2142-2144 (1995).

[0201] The examples set forth below are intended to illustrate but not limit the invention.

Examples

[0202] In the following studies a group of subjects was selected according to specific parameters relating to osteoarthritis. Nucleic acid samples obtained from individuals in the study group were subjected to genetic analysis, which identified associations between osteoarthritis and a polymorphism in the *ADAMTS2* gene on chromosome five. The polymorphism was genotyped again in two replication cohorts consisting of individuals selected for OA. In addition, SNPs proximal to the incident polymorphism were identified and allelotyped in OA case and control pools. Methods are described for producing *ADAMTS2* polypeptide and *ADAMTS2* polypeptide variants *in vitro* or *in vivo*, *ADAMTS2* nucleic acids or polypeptides and variants thereof are utilized for screening test molecules for those that interact with *ADAMTS2* molecules. Test molecules identified as interactors with *ADAMTS2* molecules and *ADAMTS2* variants are further screened *in vivo* to determine whether they treat osteoarthritis.

Example 1

Samples and Pooling Strategies

Sample Selection

[0203] Blood samples were collected from individuals diagnosed with knee osteoarthritis, which were referred to as case samples. Also, blood samples were collected from individuals not diagnosed with knee osteoarthritis as gender and age-matched controls. A database was created that listed all phenotypic trait information gathered from individuals for each case and control sample. Genomic DNA was extracted from each of the blood samples for genetic analyses.

DNA Extraction from Blood Samples

[0204] Six to ten milliliters of whole blood was transferred to a 50 ml tube containing 27 ml of red cell lysis solution (RCL). The tube was inverted until the contents were mixed. Each tube was incubated for 10 minutes at room temperature and inverted once during the incubation. The tubes were then centrifuged for 20 minutes at 3000 x g and the supernatant was carefully poured off. 100-200 μ l of residual liquid was left in the tube and was pipetted repeatedly to resuspend the pellet in the residual supernatant. White cell lysis solution (WCL) was added to the tube and pipetted repeatedly until

completely mixed. While no incubation was normally required, the solution was incubated at 37°C or room temperature if cell clumps were visible after mixing until the solution was homogeneous. 2 ml of protein precipitation was added to the cell lysate. The mixtures were vortexed vigorously at high speed for 20 sec to mix the protein precipitation solution uniformly with the cell lysate, and then centrifuged for 10 minutes at 3000 x g. The supernatant containing the DNA was then poured into a clean 15 ml tube, which contained 7 ml of 100% isopropanol. The samples were mixed by inverting the tubes gently until white threads of DNA were visible. Samples were centrifuged for 3 minutes at 2000 x g and the DNA was visible as a small white pellet. The supernatant was decanted and 5 ml of 70% ethanol was added to each tube. Each tube was inverted several times to wash the DNA pellet, and then centrifuged for 1 minute at 2000 x g. The ethanol was decanted and each tube was drained on clean absorbent paper. The DNA was dried in the tube by inversion for 10 minutes, and then 1000 µl of 1X TE was added. The size of each sample was estimated, and less TE buffer was added during the following DNA hydration step if the sample was smaller. The DNA was allowed to rehydrate overnight at room temperature, and DNA samples were stored at 2-8°C.

[0205] DNA was quantified by placing samples on a hematology mixer for at least 1 hour. DNA was serially diluted (typically 1:80, 1:160, 1:320, and 1:640 dilutions) so that it would be within the measurable range of standards. 125 µl of diluted DNA was transferred to a clear U-bottom microtitre plate, and 125 µl of 1X TE buffer was transferred into each well using a multichannel pipette. The DNA and 1X TE were mixed by repeated pipetting at least 15 times, and then the plates were sealed. 50 µl of diluted DNA was added to wells A5-H12 of a black flat bottom microtitre plate. Standards were inverted six times to mix them, and then 50 µl of 1X TE buffer was pipetted into well A1, 1000 ng/ml of standard was pipetted into well A2, 500 ng/ml of standard was pipetted into well A3, and 250 ng/ml of standard was pipetted into well A4. PicoGreen (Molecular Probes, Eugene, Oregon) was thawed and freshly diluted 1:200 according to the number of plates that were being measured. PicoGreen was vortexed and then 50µl was pipetted into all wells of the black plate with the diluted DNA. DNA and PicoGreen were mixed by pipetting repeatedly at least 10 times with the multichannel pipette. The plate was placed into a Fluoroskan Ascent Machine (microplate fluorometer produced by Labsystems) and the samples were allowed to incubate for 3 minutes before the machine was run using filter pairs 485 nm excitation and 538 nm emission wavelengths. Samples having measured DNA concentrations of greater than 450 ng/µl were re-measured for conformation. Samples having measured DNA concentrations of 20 ng/µl or less were re-measured for confirmation.

Pooling Strategies – Discovery Cohort

[0206] Samples were derived from the Nottingham knee OA family study (UK) where index cases were identified through a knee replacement registry. Siblings were approached and assessed with knee x-rays and assigned status as affected or unaffected. In all 1,157 individuals were available. In order to create same-sex pools of appropriate sizes, 335 unrelated female individuals with OA from the Nottingham OA sample were selected for the case pool. The control pool was made up of unrelated female individuals from the St. Thomas twin study (England) with normal knee x-rays and without other indications of OA, regardless of anatomical location, as well as lacking family history of OA. The St. Thomas twin study consists of Caucasian, female participants from the St. Thomas' Hospital, London, adult-twin registry, which is a voluntary registry of >4,000 twin pairs ranging from 18 to 76 years of age. The female case samples and female control samples are described further in Table 1 below.

[0207] A select set of samples from each group were utilized to generate pools, and one pool was created for each group. Each individual sample in a pool was represented by an equal amount of genomic DNA. For example, where 25 ng of genomic DNA was utilized in each PCR reaction and there were 200 individuals in each pool, each individual would provide 125 pg of genomic DNA. Inclusion or exclusion of samples for a pool was based upon the following criteria: the sample was derived from an individual characterized as Caucasian; the sample was derived from an individual of British paternal and maternal descent; case samples were derived from individuals diagnosed with specific knee osteoarthritis (OA) and were recruited from an OA knee replacement clinic. Control samples were derived from individuals free of OA, family history of OA, and rheumatoid arthritis. Also, sufficient genomic DNA was extracted from each blood sample for all allelotyping and genotyping reactions performed during the study. Phenotype information from each individual was collected and included age of the individual, gender, family history of OA, general medical information (e.g., height, weight, thyroid disease, diabetes, psoriasis, hysterectomy), joint history (previous and current symptoms, joint-related operations, age at onset of symptoms, date of primary diagnosis, age of individual as of primary diagnosis and order of involvement), and knee-related findings (crepitus, restricted passive movement, bony swelling/deformity). Additional knee information included knee history, current symptoms, any major knee injury, meniscectomy, knee replacement surgery, age of surgery, and treatment history (including hormone replace therapy (HRT)). Samples that met these criteria were added to appropriate pools based on disease status.

[0208] The selection process yielded the pools set forth in Table 1, which were used in the studies that follow:

TABLE 1

	Female case	Female control
Pool size (Number)	335	335
Pool Criteria (ex: case/control)	control	case
Mean Age (ex: years)	57.21	69.95

Example 2

Association of Polymorphic Variants with Osteoarthritis

[0209] A whole-genome screen was performed to identify particular SNPs associated with occurrence of osteoarthritis. As described in Example 1, two sets of samples were utilized, which included samples from female individuals having knee osteoarthritis (osteoarthritis cases), and samples from female individuals not having knee osteoarthritis (female controls). The initial screen of each pool was performed in an allelotyping study, in which certain samples in each group were pooled. By pooling DNA from each group, an allele frequency for each SNP in each group was calculated. These allele frequencies were then compared to one another. Particular SNPs were considered as being associated with osteoarthritis when allele frequency differences calculated between case and control pools were statistically significant. SNP disease association results obtained from the allelotyping study were then validated by genotyping each associated SNP across all samples from each pool. The results of the genotyping then were analyzed, allele frequencies for each group were calculated from the individual genotyping results, and a p-value was calculated to determine whether the case and control groups had statistically significant differences in allele frequencies for a particular SNP. When the genotyping results agreed with the original allelotyping results, the SNP disease association was considered validated at the genetic level.

SNP Panel Used for Genetic Analyses

[0210] A whole-genome SNP screen began with an initial screen of approximately 25,000 SNPs over each set of disease and control samples using a pooling approach. The pools studied in the screen are described in Example 1. The SNPs analyzed in this study were part of a set of 25,488 SNPs confirmed as being statistically polymorphic as each is characterized as having a minor allele frequency of greater than 10%. The SNPs in the set reside in genes or in close proximity to genes, and many reside

in gene exons. Specifically, SNPs in the set are located in exons, introns, and within 5,000 base-pairs upstream of a transcription start site of a gene. In addition, SNPs were selected according to the following criteria: they are located in ESTs; they are located in Locuslink or Ensembl genes; and they are located in Genomatix promoter predictions. SNPs in the set were also selected on the basis of even spacing across the genome, as depicted in Table 2.

[0211] A case-control study design using a whole genome association strategy involving approximately 28,000 single nucleotide polymorphisms (SNPs) was employed. Approximately 25,000 SNPs were evenly spaced in gene-based regions of the human genome with a median inter-marker distance of about 40,000 base pairs. Additionally, approximately 3,000 SNPs causing amino acid substitutions in genes described in the literature as candidates for various diseases were used. The case-control study samples were of female Caucasian origin (British paternal and maternal descent) 670 individuals were equally distributed in two groups: female controls and female cases. The whole genome association approach was first conducted on 2 DNA pools representing the 2 groups. Significant markers were confirmed by individual genotyping.

TABLE 2

<u>General Statistics</u>		<u>Spacing Statistics</u>	
Total # of SNPs	25,488	Median	37,058 bp
# of Exonic SNPs	>4,335 (17%)	Minimum*	1,000 bp
# SNPs with refSNP ID	20,776 (81%)	Maximum*	3,000,000 bp
Gene Coverage	>10,000	Mean	122,412 bp
Chromosome Coverage	All	Std Deviation	373,325 bp
		<i>*Excludes outliers</i>	

Allelotyping and Genotyping Results

[0212] The genetic studies summarized above and described in more detail below identified an allelic variant in the ADAMTS2 gene that is associated with osteoarthritis.

Assay for Verifying, Allelotyping, and Genotyping SNPs

[0213] A MassARRAY™ system (Sequenom, Inc.) was utilized to perform SNP genotyping in a high-throughput fashion. This genotyping platform was complemented by a homogeneous, single-tube assay method (hMET™ or homogeneous MassEXTEND™ (Sequenom, Inc.)) in which two genotyping primers anneal to and amplify a genomic target surrounding a polymorphic site of interest. A third primer (the MassEXTEND™ primer), which is complementary to the amplified target up to but not

including the polymorphism, was then enzymatically extended one or a few bases through the polymorphic site and then terminated.

[0214] For each polymorphism, SpectroDESIGNER™ software (Sequenom, Inc.) was used to generate a set of PCR primers and a MassEXTEND™ primer which were used to genotype the polymorphism. Other primer design software could be used or one of ordinary skill in the art could manually design primers based on his or her knowledge of the relevant factors and considerations in designing such primers. Table 3 shows PCR primers and Table 4 shows extension primers used for analyzing polymorphisms. The initial PCR amplification reaction was performed in a 5 µl total volume containing 1X PCR buffer with 1.5 mM MgCl₂ (Qiagen), 200 µM each of dATP, dGTP, dCTP, dTTP (Gibco-BRL), 2.5 ng of genomic DNA, 0.1 units of HotStar DNA polymerase (Qiagen), and 200 nM each of forward and reverse PCR primers specific for the polymorphic region of interest.

TABLE 3: PCR Primers

SNP Reference	Forward PCR primer	Reverse PCR primer
rs398829	ACGTTGGATGTAGTCATCGTCCGCAGCATG	ACGTTGGATGAAGACGGTGCTCTCCTTG

[0215] Samples were incubated at 95°C for 15 minutes, followed by 45 cycles of 95°C for 20 seconds, 56°C for 30 seconds, and 72°C for 1 minute, finishing with a 3 minute final extension at 72°C. Following amplification, shrimp alkaline phosphatase (SAP) (0.3 units in a 2 µl volume) (Amersham Pharmacia) was added to each reaction (total reaction volume was 7 µl) to remove any residual dNTPs that were not consumed in the PCR step. Samples were incubated for 20 minutes at 37°C, followed by 5 minutes at 85°C to denature the SAP.

[0216] Once the SAP reaction was complete, a primer extension reaction was initiated by adding a polymorphism-specific MassEXTEND™ primer cocktail to each sample. Each MassEXTEND™ cocktail included a specific combination of dideoxynucleotides (ddNTPs) and deoxynucleotides (dNTPs) used to distinguish polymorphic alleles from one another. Methods for verifying, allelotyping and genotyping SNPs are disclosed, for example, in U.S. Pat. No. 6,258,538, the content of which is hereby incorporated by reference. In Table 4, ddNTPs are shown and the fourth nucleotide not shown is the dNTP.

TABLE 4: Extension Primers

SNP Reference	Extend Probe	Termination Mix
rs398829	TGGCGTGCTCCTAGGA	ACG

[0217] The MassEXTEND™ reaction was performed in a total volume of 9 µl, with the addition of 1X ThermoSequenase buffer, 0.576 units of ThermoSequenase (Amersham Pharmacia), 600 nM MassEXTEND™ primer, 2 mM of ddATP and/or ddCTP and/or ddGTP and/or ddTTP, and 2 mM of dATP or dCTP or dGTP or dTTP. The deoxy nucleotide (dNTP) used in the assay normally was complementary to the nucleotide at the polymorphic site in the amplicon. Samples were incubated at 94°C for 2 minutes, followed by 55 cycles of 5 seconds at 94°C, 5 seconds at 52°C, and 5 seconds at 72°C.

[0218] Following incubation, samples were desalted by adding 16 µl of water (total reaction volume was 25 µl), 3 mg of SpectroCLEAN™ sample cleaning beads (Sequenom, Inc.) and allowed to incubate for 3 minutes with rotation. Samples were then robotically dispensed using a piezoelectric dispensing device (SpectroJET™ (Sequenom, Inc.)) onto either 96-spot or 384-spot silicon chips containing a matrix that crystallized each sample (SpectroCHIP™ (Sequenom, Inc.)). Subsequently, MALDI-TOF mass spectrometry (Biflex and Autoflex MALDI-TOF mass spectrometers (Bruker Daltonics) can be used) and SpectroTYPER RT™ software (Sequenom, Inc.) were used to analyze and interpret the SNP genotype for each sample.

Genetic Analysis

[0219] Minor allelic frequencies for the polymorphisms set forth in Table A were verified as being 10% or greater using the extension assay described above in a group of samples isolated from 92 individuals originating from the state of Utah in the United States, Venezuela and France (Coriell cell repositories).

[0220] Genotyping results are shown for female pools in Table 5. In Table 5, "AF" refers to allelic frequency; and "F case" and "F control" refer to female case and female control groups, respectively.

TABLE 5: Genotyping Results

SNP Reference	AF F case	AF F control	p-value
rs398829	G = 0.740 A = 0.260	G = 0.652 A = 0.348	0.0002

[0221] All of the single marker alleles set forth in Table A were considered validated, since the genotyping data agreed with the allelotyping data and each SNP significantly associated with osteoarthritis. Particularly significant associations with osteoarthritis are indicated by a calculated p-value of less than 0.05 for genotyping results.

Example 3

Association of Polymorphic Variants with Osteoarthritis in Replication Cohorts

[0222] The single marker polymorphism set forth in Table A was genotyped again in two replication cohorts consisting of individuals selected for OA.

Sample Selection and Pooling Strategies – Replication Sample 1

[0223] A second case control sample (replication sample #1) was created by using 100 Caucasian female cases from Chingford, UK, and 148 unrelated female cases from the St. Thomas twin study. Cases were defined as having Kellgren-Lawrence (KL) scores of at least 2 in at least one knee x-ray. In addition, 199 male knee replacement cases from Nottingham were included. (For a cohort description, see the Nottingham description provided in Example 1). The control pool was made up of unrelated female individuals from the St. Thomas twin study (England) with normal knee x-rays and without other indications of OA, regardless of anatomical location, as well as lacking family history of OA. The St. Thomas twin study consists of Caucasian, female participants from the St. Thomas' Hospital, London, adult-twin registry, which is a voluntary registry of >4,000 twin pairs ranging from 18 to 76 years of age. The replication sample 1 cohort was used to replicate the initial results. Table 6 below summarizes the selected phenotype data collected from the case and control individuals.

TABLE 6

Phenotype	Female cases (n=248): median (range)/ (n,%)	Male cases (n=199): median (range)/ (n,%)	Female controls (n=313): mean (range)/ (n,%)
Age	59 (39- 73)	66 (45- 73)	55 (50- 72)
Height (cm)	162 (141- 178)	175 (152- 198)	162 (141- 176)
Weight (kg)	68 (51- 123)	86 (62- 127)	64 (40- 111)
Body mass index (kg/m ²)	26 (18- 44)	29 (21- 41)	24 (18- 46)
Kellgren- Lawrence* left knee	0 (63, 26%), 1 (20, 8%), 2 (105, 43%), 3 (58, 23%), 4 (1, 0%)	NA	NA
Kellgren- Lawrence* right knee	0 (43, 7%), 1 (18, 7%), 2 (127, 52%), 3 (57, 23%), 4 (1, 0%)	NA	NA
KL* >2 both knees	No (145, 59%), Yes (101, 41%)	NA	NA
KL* >2 either knee	No (0, 0%), Yes (248, 100%)	NA	NA

* 0: normal, 1: doubtful, 2: definite osteophyte (bony protuberance), 3: joint space narrowing (with or without osteophyte), 4: joint deformity

Sample Selection and Pooling Strategies – Replication Sample 2

[0224] A third case control sample (replication sample #2) was created by using individuals with symptoms of OA from Newfoundland, Canada. These individuals were recruited and examined by rheumatologists. Affected joints were x-rayed and a final diagnosis of definite or probable OA was made according to American College of Rheumatology criteria by a single rheumatologist to avoid any inter-examiner diagnosis variability. Controls were recruited from volunteers without any symptoms from the musculoskeletal system based on a normal joint exam performed by a rheumatologist. Only cases with a diagnosis of definite OA were included in the study. Only individuals of Caucasian origin were included. The cases consisted of 228 individuals with definite knee OA, 106 individuals with definite hip OA, and 74 individuals with hip OA.

TABLE 7

Phenotype	Case	Control
Age at Visit	62.7	52.5
Sex (Female/Male)	227/119	174/101
Knee OA Xray: No	35% (120)	80% (16)
Unknown	1% (4)	0% (0)
Yes	64% (221)	20% (4)
Hip OA Xray: No	63% (215)	80% (16)
Unknown	2% (7)	0% (0)
Yes	35% (121)	20% (4)

Assay for Verifying, Allelotyping, and Genotyping SNPs

[0225] Genotyping of the replication cohorts described in Tables 6 and 7 was performed using the same methods used for the original genotyping, as described herein. A MassARRAY™ system (Sequenom, Inc.) was utilized to perform SNP genotyping in a high-throughput fashion. This genotyping platform was complemented by a homogeneous, single-tube assay method (hME™ or homogeneous MassEXTEND™ (Sequenom, Inc.)) in which two genotyping primers anneal to and amplify a genomic target surrounding a polymorphic site of interest. A third primer (the MassEXTEND™ primer), which is complementary to the amplified target up to but not including the polymorphism, was then enzymatically extended one or a few bases through the polymorphic site and then terminated.

[0226] For each polymorphism, SpectroDESIGNER™ software (Sequenom, Inc.) was used to generate a set of PCR primers and a MassEXTEND™ primer which were used to genotype the polymorphism. Other primer design software could be used or one of ordinary skill in the art could

manually design primers based on his or her knowledge of the relevant factors and considerations in designing such primers. Table 3 shows PCR primers and Table 4 shows extension probes used for analyzing (e.g., genotyping) polymorphisms in the replication cohorts. The initial PCR amplification reaction was performed in a 5 μ l total volume containing 1X PCR buffer with 1.5 mM $MgCl_2$ (Qiagen), 200 μ M each of dATP, dGTP, dCTP, dTTP (Gibco-BRL), 2.5 ng of genomic DNA, 0.1 units of HotStar DNA polymerase (Qiagen), and 200 nM each of forward and reverse PCR primers specific for the polymorphic region of interest.

[0227] Samples were incubated at 95°C for 15 minutes, followed by 45 cycles of 95°C for 20 seconds, 56°C for 30 seconds, and 72°C for 1 minute, finishing with a 3 minute final extension at 72°C. Following amplification, shrimp alkaline phosphatase (SAP) (0.3 units in a 2 μ l volume) (Amersham Pharmacia) was added to each reaction (total reaction volume was 7 μ l) to remove any residual dNTPs that were not consumed in the PCR step. Samples were incubated for 20 minutes at 37°C, followed by 5 minutes at 85°C to denature the SAP.

[0228] Once the SAP reaction was complete, a primer extension reaction was initiated by adding a polymorphism-specific MassEXTEND™ primer cocktail to each sample. Each MassEXTEND™ cocktail included a specific combination of dideoxynucleotides (ddNTPs) and deoxynucleotides (dNTPs) used to distinguish polymorphic alleles from one another. Methods for verifying, allelotyping and genotyping SNPs are disclosed, for example, in U.S. Pat. No. 6,258,538, the content of which is hereby incorporated by reference. In Table 7, ddNTPs are shown and the fourth nucleotide not shown is the dNTP.

[0229] The MassEXTEND™ reaction was performed in a total volume of 9 μ l, with the addition of 1X ThermoSequenase buffer, 0.576 units of ThermoSequenase (Amersham Pharmacia), 600 nM MassEXTEND™ primer, 2 mM of ddATP and/or ddCTP and/or ddGTP and/or ddTTP, and 2 mM of dATP or dCTP or dGTP or dTTP. The deoxy nucleotide (dNTP) used in the assay normally was complementary to the nucleotide at the polymorphic site in the amplicon. Samples were incubated at 94°C for 2 minutes, followed by 55 cycles of 5 seconds at 94°C, 5 seconds at 52°C, and 5 seconds at 72°C.

[0230] Following incubation, samples were desalted by adding 16 μ l of water (total reaction volume was 25 μ l), 3 mg of SpectroCLEAN™ sample cleaning beads (Sequenom, Inc.) and allowed to incubate for 3 minutes with rotation. Samples were then robotically dispensed using a piezoelectric dispensing device (SpectroJET™ (Sequenom, Inc.)) onto either 96-spot or 384-spot silicon chips containing a matrix that crystallized each sample (SpectroCHIP™ (Sequenom, Inc.)). Subsequently, MALDI-TOF mass spectrometry (Biflex and Autoflex MALDI-TOF mass spectrometers (Bruker Daltonics) can be used) and

SpectroTYPER RT™ software (Sequenom, Inc.) were used to analyze and interpret the SNP genotype for each sample.

Genetic Analysis

[0231] Genotyping results for replication cohorts #1 and #2 are provided in Tables 8 and 9, respectively.

TABLE 8

rsID	Replication #1 (Mixed Male/Female cases and Female controls)				Meta-analysis Disc. + Rep #1
	AF OA Con	AF OA Cas	Delta	P-value	P-value
rs398829	0.30	0.28	0.02	0.307	0.0260

TABLE 9

rsID	Replication #2 (Newfoundland) (Male/Female cases and controls)				Meta-analysis Disc. + Rep #2
	AF OA Con	AF OA Cas	Delta	P-value	Not Done
rs398829	0.27	0.28	-0.013	0.627	

[0232] To combine the evidence for association from multiple sample collections, a meta-analysis procedure was employed. The allele frequencies were compared between cases and controls within the discovery sample, as well as within the replication cohort #1 using the DerSimonian-Laird approach (DerSimonian, R. and N. Laird. 1986. Meta-analysis in clinical trials. Control Clin Trials 7: 177-188.)

[0233] The absence of a statistically significant association in one or more of the replication cohorts should not be interpreted as minimizing the value of the original finding. There are many reasons why a biologically derived association identified in a sample from one population would not replicate in a sample from another population. The most important reason is differences in population history. Due to bottlenecks and founder effects, there may be common disease predisposing alleles present in one population that are relatively rare in another, leading to a lack of association in the candidate region. Also, because common diseases such as arthritis-related disorders are the result of susceptibilities in many genes and many environmental risk factors, differences in population-specific genetic and environmental backgrounds could mask the effects of a biologically relevant allele. For these and other reasons, statistically strong results in the original, discovery sample that did not replicate in one or more of the replication samples may be further evaluated in additional replication cohorts and experimental systems.

Example 4

ADAMTS2 Region Proximal SNPs

[0234] It has been discovered that SNP rs398829 in *ADAMTS2* is associated with occurrence of osteoarthritis in subjects. This gene encodes a disintegrin and metalloproteinase with thrombospondin motifs-2 (*ADAMTS2*), which is a member of the *ADAMTS* protein family. Members of the family share several distinct protein modules, including a propeptide region, a metalloproteinase domain, a disintegrin-like domain, and a thrombospondin type 1 (TS) motif. *ADAMTS2* is involved in collagens 1, 2 and 5 N-terminal processing, (type II collagen is the major form in cartilage). Mutations in this gene cause Ehlers-Danlos syndrome type VIIC, a recessively inherited connective-tissue disorder that causes loose joints and fragile skin. Mild loss of function may exacerbate physical joint damage leading to a predisposition to OA and incorrectly processed collagen can act dominantly to inhibit self assembly of fibrils. Alternative splicing of the gene generates 2 transcript variants. The short transcript encodes a protein, which has no significant procollagen N-peptidase activity.

[0235] Two hundred-nine additional allelic variants proximal to rs398829 were identified and subsequently allelotyped in osteoarthritis case and control sample sets as described in Examples 1 and 2. The polymorphic variants are set forth in Table 10. The chromosome positions provided in column four of Table 10 are based on Genome "Build 34" of NCBI's GenBank.

TABLE 10

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs2278221	5	210	178695460	c/t
rs1650358	5	3608	178698858	c/g
rs1643818	5	3609	178698859	c/g
rs3733916	5	4318	178699568	c/t
rs1624933	5	5593	178700843	a/g
rs1624857	5	5629	178700879	c/t
rs1624832	5	5639	178700889	a/g
rs1624829	5	5640	178700890	c/t
rs2161171	5	8943	178704193	a/c
rs1530499	5	17968	178713218	a/g
rs888764	5	19887	178715137	a/g
rs873987	5	21034	178716284	a/g
rs4078699	5	21085	178716335	c/t
rs870311	5	21596	178716846	a/g
rs1643817	5	23379	178718629	a/c
rs1643816	5	23432	178718682	a/c
rs1650355	5	24007	178719257	a/c
rs888763	5	26121	178721371	a/g
rs1862212	5	26273	178721523	a/t

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs1110514	5	26755	178722005	a/t
rs3797600	5	27411	178722661	c/t
rs3797602	5	27710	178722960	g/t
rs3797603	5	27842	178723092	c/t
rs3776819	5	28379	178723629	c/t
rs252076	5	29603	178724853	c/t
rs252075	5	31232	178726482	c/g
rs252074	5	31504	178726754	a/g
rs252068	5	32583	178727833	c/g
rs252069	5	32794	178728044	a/g
rs194040	5	32840	178728090	c/t
rs252070	5	33044	178728294	c/t
rs3797606	5	33150	178728400	a/c
rs171667	5	33218	178728468	a/g
rs187539	5	33513	178728763	c/t
rs3836834	5	33959	178729209	- /tatcaaacctacatga aa
rs252071	5	34486	178729736	a/g
rs252072	5	36289	178731539	c/t
rs252073	5	36570	178731820	c/t
rs379589	5	38247	178733497	a/t
rs2052472	5	38477	178733727	a/c
rs2052471	5	38518	178733768	c/t
rs2052470	5	38529	178733779	c/t
rs2052469	5	38667	178733917	a/g
rs3797608	5	39781	178735031	c/t
rs3797609	5	39856	178735106	c/t
rs3822601	5	39927	178735177	c/t
rs153131	5	40506	178735756	a/g
rs751546	5	41869	178737119	c/g
rs2279979	5	42452	178737702	c/t
rs252060	5	44788	178740038	c/t
rs3797610	5	46059	178741309	a/c
rs194039	5	46846	178742096	a/g
rs168773	5	47712	178742962	a/t
rs252061	5	48796	178744046	c/t
rs187537	5	49441	178744691	c/g
rs252062	5	49602	178744852	a/t
rs2431255	5	49723	178744973	a/c
rs3797612	5	50050	178745300	c/t
rs3797613	5	50171	178745421	c/t
rs614114	5	50477	178745727	c/t
rs252063	5	50818	178746068	c/t
rs252064	5	50833	178746083	c/t
rs252065	5	50881	178746131	a/g
rs450502	5	50882	178746132	a/g
rs439252	5	51386	178746636	c/t

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs252066	5	51534	178746784	c/t
rs457957	5	52317	178747567	a/g
rs3797614	5	52368	178747618	c/t
rs423552	5	52970	178748220	a/g
rs398829	5	53023	178748273	a/g
rs416646	5	53356	178748606	a/g
rs187450	5	53882	178749132	g/t
rs337807	5	54553	178749803	c/t
rs337806	5	55475	178750725	a/c
rs1396438	5	55530	178750780	a/g
rs1396437	5	55691	178750941	c/t
rs2411811	5	55848	178751098	a/c
rs2898813	5	55879	178751129	c/g
rs189256	5	56316	178751566	a/g
rs173072	5	56911	178752161	a/c
rs337805	5	57320	178752570	a/g
rs191415	5	57391	178752641	c/t
rs180045	5	57437	178752687	c/t
rs189255	5	57478	178752728	c/g
rs652766	5	57500	178752750	c/t
rs466750	5	59111	178754361	g/t
rs442406	5	59333	178754583	a/g
rs662407	5	59715	178754965	a/g
rs592971	5	59804	178755054	a/g
rs457187	5	59851	178755101	a/g
rs459490	5	59929	178755179	c/t
rs459668	5	60052	178755302	c/t
rs462646	5	60240	178755490	c/t
rs458272	5	60359	178755609	g/t
rs463455	5	60381	178755631	a/g
rs675880	5	60456	178755706	c/t
rs810617	5	60724	178755974	c/g
rs464156	5	60875	178756125	c/t
rs458083	5	60968	178756218	a/g
rs467333	5	60978	178756228	c/g
rs465381	5	60998	178756248	c/t
rs466363	5	61557	178756807	c/t
rs2457099	5	62091	178757341	c/t
rs463901	5	62645	178757895	c/t
rs465621	5	62943	178758193	a/c
rs463724	5	63131	178758381	a/t
rs465242	5	63145	178758395	g/t
rs467419	5	63406	178758656	a/g
rs456135	5	63427	178758677	c/g
rs464536	5	63554	178758804	c/t
rs461898	5	63661	178758911	a/g
rs389558	5	64093	178759343	a/g

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs466752	5	64153	178759403	c/t
rs455655	5	64409	178759659	c/g
rs463435	5	64544	178759794	c/t
rs2174971	5	65257	178760507	c/t
rs1979979	5	65626	178760876	a/g
rs411804	5	65739	178760989	a/g
rs1623885	5	66392	178761642	c/t
rs1643811	5	66720	178761970	c/t
rs434430	5	69177	178764427	a/t
rs187538	5	69336	178764586	g/t
rs252067	5	69636	178764886	a/g
rs459319	5	69823	178765073	a/g
rs467289	5	69928	178765178	c/t
rs462644	5	70547	178765797	c/t
rs458752	5	70633	178765883	c/t
rs708320	5	71805	178767055	a/c
rs457954	5	72181	178767431	c/g
rs2411810	5	72200	178767450	c/t
rs3084687	5	72474	178767724	-/at
rs69638	5	72567	178767817	c/g
rs455452	5	72973	178768223	a/g
rs464850	5	73468	178768718	a/g
rs431472	5	73889	178769139	a/g
rs2411809	5	75730	178770980	c/t
rs2457094	5	75970	178771220	a/g
rs2457095	5	76114	178771364	a/g
rs2261740	5	76342	178771592	c/t
rs1109180	5	76449	178771699	a/g
rs1109179	5	76465	178771715	c/t
rs1109178	5	76791	178772041	a/c
rs456909	5	78042	178773292	a/g
rs469124	5	80758	178776008	a/g
rs468039	5	80778	178776028	c/t
rs467017	5	81356	178776606	a/c
rs469290	5	81576	178776826	a/g
rs469090	5	81689	178776939	c/t
rs469568	5	81759	178777009	g/t
rs468386	5	81950	178777200	c/g
rs469349	5	82562	178777812	a/c
rs469099	5	83591	178778841	c/t
rs456868	5	83700	178778950	a/g
rs465389	5	83821	178779071	c/g
rs463892	5	83842	178779092	c/g
rs468548	5	83923	178779173	g/t
rs654612	5	83929	178779179	a/c
rs468542	5	84021	178779271	c/g
rs469262	5	84175	178779425	c/t

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs708323	5	84417	178779667	a/g
rs469089	5	84747	178779997	c/g
rs469396	5	85746	178780996	c/g
rs468723	5	86129	178781379	c/t
rs467604	5	86335	178781585	a/g
rs338874	5	87315	178782565	c/g
rs338875	5	87648	178782898	a/g
rs1385803	5	87764	178783014	a/c
rs1385804	5	87770	178783020	c/g
rs338876	5	88221	178783471	c/t
rs189803	5	90474	178785724	a/c
rs452215	5	91148	178786398	g/t
rs641170	5	91150	178786400	g/t
rs584398	5	91160	178786410	g/t
rs385330	5	91733	178786983	c/t
rs429538	5	91772	178787022	a/c
rs371229	5	91785	178787035	c/t
rs460874	5	93140	178788390	a/t
rs646121	5	93148	178788398	a/t
rs468262	5	96080	178791330	a/g
rs467863	5	96157	178791407	c/g
rs191434	5	96313	178791563	a/c
rs2054782	5	96759	178792009	c/t
rs468499	5	97026	178792276	a/c
rs180287	5	97320	178792570	c/g
rs338877	5	97732	178792982	a/t
rs650665	5	98713	178793963	c/g
rs193419	5	99707	178794957	a/c
rs180288	5	99959	178795209	c/g
rs186834	5	100009	178795259	a/g
rs189266	5	100020	178795270	c/g
rs189267	5	100065	178795315	a/c
rs170937	5	100086	178795336	c/g
rs463263	5	101270	178796520	c/g
rs463262	5	101276	178796526	g/t
rs460454	5	101371	178796621	c/t
rs460455	5	101376	178796626	c/g
rs460505	5	101439	178796689	c/t
rs931316	5	101820	178797070	c/t
rs463431	5	102392	178797642	c/g
rs461542	5	102602	178797852	a/g
rs463557	5	102604	178797854	a/c
rs191453	5	102896	178798146	c/t
rs2271212	5	189104	178884354	c/t
rs462009	5	189134	178884384	c/t
rs2271211	5	189205	178884455	a/g
rs396474	5	Not mapped	Not mapped	a/c

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs428901	5	Not mapped	Not mapped	a/t
rs452300	5	Not mapped	Not mapped	g/t
rs670256	5	Not mapped	Not mapped	g/t

Assay for Verifying and Allelotyping SNPs

[0236] The methods used to verify and allelotype the 209 proximal SNPs of Table 10 are the same methods described in Examples 1 and 2 herein. The primers and probes used in these assays are provided in Table 11 and Table 12, respectively.

TABLE 11

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs2278221	ACGTTGGATGTCATGGGCCACCACAAAC	ACGTTGGATGTATGCTCCTGTACCCGGCAT
rs1650358	ACGTTGGATGTGGATGGCTCCATGTTCTTG	ACGTTGGATGAAGTGCTGGGATACAGGTG
rs1643818	ACGTTGGATGCTGGGATTACAGGTGTGAAC	ACGTTGGATGTGGATGGCTCCATGTTCTTG
rs3733916	ACGTTGGATGCCGAGCAGGCTGTAGTGTG	ACGTTGGATGCTTTGTACCACTGGAACAG
rs1624933	ACGTTGGATGAGGCTGGTCTCAAACCTCTG	ACGTTGGATGAACAAAAGTTGGCCGTGC
rs1624857	ACGTTGGATGTGAGGTGAGGAGTTTGAGAC	ACGTTGGATGGCCCAAGCCAGACTAAGT
rs1624832	ACGTTGGATGTGAGGTGAGGAGTTTGAGAC	ACGTTGGATGGCCCAAGCCAGACTAAGT
rs1624829	ACGTTGGATGTGAGGTGAGGAGTTTGAGAC	ACGTTGGATGGCCCAAGCCAGACTAAGT
rs2161171	ACGTTGGATGCCCGTCAACCATTTAATTC	ACGTTGGATGAGAGTGGATCCAGTCTGCAG
rs1530499	ACGTTGGATGACTCCAAGATTTCCCATTTT	ACGTTGGATGTTCTGTGTTCCACCTATGG
rs888764	ACGTTGGATGTAGTTGAATGTTGTATTGGC	ACGTTGGATGACCGTGATAAACACAGAATG
rs873987	ACGTTGGATGGCTGTTAATCATGTGTCGGG	ACGTTGGATGATTGGCCCATCACCAGAC
rs4078699	ACGTTGGATGGTACCGTGGATTCTTTTAGG	ACGTTGGATGGTATTGGAAAAGAGCAGAGAC
rs870311	ACGTTGGATGTGAGGCTCCAGTGTGTAAG	ACGTTGGATGAAAAGGAGGAGTGCCCTGTG
rs1643817	ACGTTGGATGATGGGAACTCCTGGTCCTG	ACGTTGGATGAAAATGCAAGCCCGCACTTG
rs1643816	ACGTTGGATGTTTTCTCCCTTTCTAGCCC	ACGTTGGATGTTGGCATGAGAGATGGACAG
rs1650355	ACGTTGGATGTCAACAGCAACAAAACCAAA	ACGTTGGATGTTAAATAGGTGACAGGGTTG
rs888763	ACGTTGGATGAAGAGGAAGAGACATACCA	ACGTTGGATGAACAACATGGACTCAGGCTG
rs1862212	ACGTTGGATGGGCCACATTTTAAACAAAGGG	ACGTTGGATGTCCTCTGAGGTTCCATATAAG
rs1110514	ACGTTGGATGTGCCCAACGTTCCATGTTTCA	ACGTTGGATGATCAGTGTAGCCCTCTTCTG
rs3797600	ACGTTGGATGCCTTCTGTCACTCTCTTTG	ACGTTGGATGGGAAGTGACTGCTGAGCTG
rs3797602	ACGTTGGATGAGAAACAGGGAATGGCTGTGT	ACGTTGGATGGCAGGCTCCGGGAAGTATG
rs3797603	ACGTTGGATGCACCATCCATCATGATGTC	ACGTTGGATGTGCTACCTCAAACAGTGGG
rs3776819	ACGTTGGATGCAAGCAACCATTCATTGCAC	ACGTTGGATGAAATGAGGATGTGCAGTCCC
rs252076	ACGTTGGATGACTTCTGACTTCAGGTGATC	ACGTTGGATGTATAGGAACGAAAGAAAGCC
rs252075	ACGTTGGATGTGGGAGCATTTGCAAGGCATG	ACGTTGGATGAAGCCCTCAGATGTTCCGGAG
rs252074	ACGTTGGATGTTGCGGATGGCTCCTGGCT	ACGTTGGATGAAGTTGAGGCTCCGGAGCA
rs252068	ACGTTGGATGGGGTAGGAAGGGTTAAGC	ACGTTGGATGGCAGCCCTCAATTCTTTAG
rs252069	ACGTTGGATGTGCCCATTTCTGTTATTCC	ACGTTGGATGTTTGGACTTGGCGTGCACCT
rs194040	ACGTTGGATGTGCCCATTTCTGTTATTCC	ACGTTGGATGTTTGGACTTGGCGTGCACCT
rs252070	ACGTTGGATGCTCAAGGACATTTGCCCTGG	ACGTTGGATGGGAGAAGCAGCTCTCCTTTC

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs3797606	ACGTTGGATGGTTTCCCCAACAGAGAGC	ACGTTGGATGGGAAATGTTCAAAGCCGACG
rs171867	ACGTTGGATGGGGAAACACATTGTAATGCG	ACGTTGGATGCGCTTCCTCATTTGCTATTCC
rs187539	ACGTTGGATGAGCCACCCCAACCTTCAGGA	ACGTTGGATGTGTCTCTGGACATGGTTTT
rs3836834	ACGTTGGATGAGTAAACCTGACTCTTGCTC	ACGTTGGATGTAGTAATCTGATCCTCGGCC
rs252071	ACGTTGGATGGGCTCAACCTGAAACAACCC	ACGTTGGATGGGGATATCCCATCTCTGAG
rs252072	ACGTTGGATGTGTGTTTCCCCAAAGCGACG	ACGTTGGATGTGTGTTTCCAGAGCTGGAG
rs252073	ACGTTGGATGGGGAAAGGCCGAGAAAAGTC	ACGTTGGATGACAAGCTCAGCAGAGTTCCTCA
rs379589	ACGTTGGATGAAACACGGGAGTACTGAGCA	ACGTTGGATGTTGTAGCTGCTCTGCCGTC
rs2052472	ACGTTGGATGAACAGCTCAAGGATCACCC	ACGTTGGATGAAAGGAGACGGTCAAGCTGTC
rs2052471	ACGTTGGATGACAGCTGACCGTCTCCTTTG	ACGTTGGATGCCCGTCTGGACAAGCTTTT
rs2052470	ACGTTGGATGACAGCTGACCGTCTCCTTTG	ACGTTGGATGCCCGTCTGGACAAGCTTTT
rs2052469	ACGTTGGATGAGGGAAAAGATATCGACGCG	ACGTTGGATGAGTGAACAACCTCTCGCCTC
rs3797808	ACGTTGGATGTGCTTTGCTTGGCTTCTGCG	ACGTTGGATGTGCACTAAGGGAGTGAGTGG
rs3797809	ACGTTGGATGTGCGAAGCCAAAGGCAAGC	ACGTTGGATGACAGCAATTTGGAGTCCCCTG
rs3822601	ACGTTGGATGAGGTCACTGAGCGCTGAGAT	ACGTTGGATGTGCTTGGCTGAAGATCGAG
rs153131	ACGTTGGATGTAATCACGTGTCTGATCCC	ACGTTGGATGAGCTGTCTCAGTCATGTTT
rs751546	ACGTTGGATGTCTCTGCTCTGCCGTTCTACA	ACGTTGGATGATCAGCTCAAAGGACCGGTG
rs2279979	ACGTTGGATGTATTTGCTACCAGGAACACGTA	ACGTTGGATGAAAAGGGGCCACTTCAGGG
rs252060	ACGTTGGATGTGGCCAGAGCCCGTGTTC	ACGTTGGATCGCGCAATCCCATCTCTAGT
rs3797810	ACGTTGGATGAAAAGCTTCTCCTTGGGTG	ACGTTGGATGCAAGTAGGGCAGAACTCAG
rs194039	ACGTTGGATGAAAGTCTGGGATTACAGGC	ACGTTGGATGTGCTGGGAGGAAGACATTAC
rs186773	ACGTTGGATGTGGTGACATGGAGATATCAC	ACGTTGGATGAGTCCCTATCCATCCTCTTC
rs252061	ACGTTGGATGTGCTCACACTCTCTTGAAG	ACGTTGGATGCTGTCTTCCCTGCTTTTTC
rs187537	ACGTTGGATGCGAGGATGTCATGCTAAGTG	ACGTTGGATGGTACCTCGCATAAAGTGGATC
rs252062	ACGTTGGATGAAGCACATTCTATGTGGCTGG	ACGTTGGATGCTGAAACTCAATGGGCACAG
rs2431255	ACGTTGGATGGGTGAAGACGGTGACTTATG	ACGTTGGATGCTGTGCTTCTTGAAGAAGTG
rs3797812	ACGTTGGATGAGTGAGGACGACAGGGCATT	ACGTTGGATGAGCGTGGCGGAGGAGATAA
rs3797813	ACGTTGGATGATCAGAGGCGAGAGCCCC	ACGTTGGATGGGCTGTCTGCAGAGGCGG
rs614114	ACGTTGGATGGGTGGAGGATGCTTAGAAC	ACGTTGGATGGGCTGGATCACTAGGGTTTG
rs252063	ACGTTGGATGTTGGAATTACAGTCCGATGG	ACGTTGGATGCTGAGAGACTGAAAAGCACA
rs252064	ACGTTGGATGCTGAGAGACTGAAAAGCACA	ACGTTGGATGTTGGAATTACAGTCCGATGG
rs252065	ACGTTGGATGAAAACCTAAGGCTCAGAGGAC	ACGTTGGATGTGGGCTTGGAAATACAGTCC
rs450502	ACGTTGGATGATGAGAAAACCAAGGCTCAG	ACGTTGGATGCTGGGCTTGAATTACAGTCT
rs439252	ACGTTGGATGATCTCTGACCTCGTGATCC	ACGTTGGATGTCATAATAACGGCCGGGTGC
rs252066	ACGTTGGATGTCTCTCTCTGACCGGCTTG	ACGTTGGATGTGTAACCAAGTAATCTGCCGATG
rs457957	ACGTTGGATGTTCACTGATCATTAGAGCGAG	ACGTTGGATGATTTCCCTCCCAATTCCTC
rs3797814	ACGTTGGATGACTGCGAGCTTTAAGGAGGG	ACGTTGGATGCCAAACAGAAACCCCTTTTC
rs423552	ACGTTGGATGGCAGGACCTCGATGTTGTAG	ACGTTGGATGATCCTAGAGGAGCAGGCCAAC
rs398829	ACGTTGGATGTAGTATCTGCTCCGACGATG	ACGTTGGATGAAGACGGTGTCTCTCTGTTG
rs416646	ACGTTGGATGGGTGGGTCTCTCACAGTCTC	ACGTTGGATGAGACAGGACACCTCTGACTT
rs187450	ACGTTGGATGAGAAGGCAGGAGCATATCC	ACGTTGGATGACCAAGATGAACCCCTCTGT
rs373807	ACGTTGGATGTACACCGATGCTGACAGCAG	ACGTTGGATGATGCTGGGATGCCATGGGTC
rs373806	ACGTTGGATGATAAAGAGATGGGGCCACC	ACGTTGGATGGCCCTGTGTGTTTGTCTCC
rs1396438	ACGTTGGATGTACCTTCTGCTGCCAGAATG	ACGTTGGATGCTGGAGACAAAACACACAG
rs1396437	ACGTTGGATGTAAAACTCTGCCTGCTCGG	ACGTTGGATGTCAGACATTTCCCTGTAGGA
rs2411811	ACGTTGGATGGAGGGATGCTCTAGAACATA	ACGTTGGATGCTGAATTTTCACTGAAATGG
rs2898813	ACGTTGGATGTCTCACCACCTTTGCTCTT	ACGTTGGATGATGCTGATAATTTTGGGGTG

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs189256	ACGTTGGATGCTCCCTATAGCAAGGCTCTA	ACGTTGGATGTAAACCCAGGCCATGAAGAG
rs173072	ACGTTGGATGAGCTGGAGATCTCTTTGCTC	ACGTTGGATGCTAAAACAGGATGGCTCTGG
rs337805	ACGTTGGATGGAACAAACCAAGGACGAGG	ACGTTGGATGATGTGGACACGTTGGACTC
rs191415	ACGTTGGATGAATTACATGACTCGGACAAG	ACGTTGGATGTGCTGGTGAAGTACAGAAAGG
rs180045	ACGTTGGATGGTCCCAAGGTTTCTGTCTC	ACGTTGGATGTGACTCTCACCAGCACTGAG
rs189255	ACGTTGGATGAGGTTCCAGACTCACTCCCA	ACGTTGGATGGGGTGATTTCGCGGAATGAG
rs652766	ACGTTGGATGGGGTGATTTCGCGGAATGAG	ACGTTGGATGACCATCCCACGATGCTCCC
rs466750	ACGTTGGATGTATCTCCTTAAATGCCTTGG	ACGTTGGATGTGACCAAGGAGGATTTAAAC
rs442406	ACGTTGGATGTGACAAGGTCACGTGTTCTG	ACGTTGGATGCCAGACAAGTCTGTATACAGC
rs662407	ACGTTGGATGCCACAGTCACCAITACTGAG	ACGTTGGATGCTTGAGCCATGAGTGGAAATG
rs592971	ACGTTGGATGGGAAGCATTTCTTTGACTGC	ACGTTGGATGATTCATCTCATGGCTCAAG
rs457187	ACGTTGGATGTGTGAGATGAGGAGTATCTG	ACGTTGGATGGCAGTCAAGAAATGCCTCC
rs459490	ACGTTGGATGACAGATACTCCTCATCTCAC	ACGTTGGATGGGGAGTTTGTCTGTATAGC
rs459668	ACGTTGGATGGCTTCATTACTGAGGTCTTC	ACGTTGGATGTGAATGTTCAACGACTACAC
rs462646	ACGTTGGATGCAATTATTCCAGCGAGATTA	ACGTTGGATGCTCCTCCAAATGAATCAAGAA
rs458272	ACGTTGGATGATGCCTCCTCATTTGTCATT	ACGTTGGATGCCAACAAAGTGATTCACAC
rs463455	ACGTTGGATGGCTCCTCCTCATTTGTCATT	ACGTTGGATGCCAACAAAGTGATTCACAC
rs675880	ACGTTGGATGCAGCTCCATTGATCTGTTTC	ACGTTGGATGAAAGTAATGACAATGAGGAGG
rs106117	ACGTTGGATGTGATCTCAGCTTACCACAGC	ACGTTGGATGATGCCTGTAAATCCCACTAC
rs464156	ACGTTGGATGCAGATCCAGAATAATGTGGG	ACGTTGGATGTTCTAGAAAGGACCAAAATC
rs458083	ACGTTGGATGTGTTGTTTCTCCCTCCTG	ACGTTGGATGTGGCTCCTTTCTAGAATCCC
rs467333	ACGTTGGATGCTTGTATTCTTCCCTCC	ACGTTGGATGTTGGCTCCTTTCTAGAATCC
rs465381	ACGTTGGATGACTTGCCTCATCTGTTCCAG	ACGTTGGATGACAAGCCTCTAAGGATAGGG
rs466363	ACGTTGGATGAAGTGACCTGAGGTGATGG	ACGTTGGATGTGAAGACAGTTTACCCCCGTG
rs2457099	ACGTTGGATGTCTCCTTACACTGCCAGCGT	ACGTTGGATGCACTGTATGTCTACTTGAGC
rs463901	ACGTTGGATGAGAGTGCCAAAGTGCAAAAGG	ACGTTGGATGTGCTTTCGCTGTGTGTACC
rs465621	ACGTTGGATGGGAAGTCATGGAAGTGCTAG	ACGTTGGATGAAAGAGCCCTAGGCTTGGAA
rs463724	ACGTTGGATGAGTGTGCCTGTCTGCCCTCA	ACGTTGGATGAAGGGGCAGATGGCACACTTG
rs465242	ACGTTGGATGTGCTGCCTGTCTGCCCTCA	ACGTTGGATGAAGGGGCAGATGGCACACTTG
rs467419	ACGTTGGATGAGTCCCAAAACGTAAGTCC	ACGTTGGATGAGTCTAATTCCTTGAGCCTC
rs456135	ACGTTGGATGAGTCTAATCCCTGAGCCTC	ACGTTGGATGACGTAAAGTCTAATGACCGC
rs464536	ACGTTGGATGTGCTCCAGGCTTTGGTCCCT	ACGTTGGATGAATTAGACATAAGGCCATGATG
rs461898	ACGTTGGATGGGGAATACACAGCCACAGAG	ACGTTGGATGAGGTCAACGGGAACAAAGGTC
rs389558	ACGTTGGATGGCACTCTGACAGTTCTCTA	ACGTTGGATGTTTCTCCCTGAAGCATGG
rs466752	ACGTTGGATGGGCCCTTCTCCTTTAGTGC	ACGTTGGATGAGTCTGACAGATTCTCTAAA
rs455655	ACGTTGGATGCTATTTCACCCCATATGCG	ACGTTGGATGAACACACAGCATCAGGTTGC
rs463435	ACGTTGGATGCTACGCCATAGCTGGATTG	ACGTTGGATGCTGTGCTGGGAAATGTGAC
rs2174971	ACGTTGGATGAACACAACCTTCCCCTTCGTC	ACGTTGGATGTGAATCCTTGGAGGTGAGTG
rs1979979	ACGTTGGATGTGGCTGTGACGACCCCACTT	ACGTTGGATGCCCAAGGAGGGAAGTAATC
rs411804	ACGTTGGATGACAGATGACAGGCGGAAATC	ACGTTGGATGAGGCTCTCCAGATGATGTCCA
rs1623885	ACGTTGGATGAATCAGCTAGGAAGAGCCTG	ACGTTGGATGTTCTGACCCCTCTAGGTCAG
rs1643811	ACGTTGGATGCAGGGCCCTGGTACTTTTCAG	ACGTTGGATGCAATGGTGGTGAATGACCTG
rs434430	ACGTTGGATGTCCAGGAGTTCACTGTAGAG	ACGTTGGATGCACATGCATACATTTCATCAG
rs187538	ACGTTGGATGACATGGGGCTTGGCAAAATG	ACGTTGGATGCACCTGCTCAGAAAGTAGCAT
rs252067	ACGTTGGATGACGAATTTGCTGTGGTGAGG	ACGTTGGATGTTTCTTGGGGAGCTGTGCGC
rs459319	ACGTTGGATGCCATCTCTGACCATAGACA	ACGTTGGATGGCTTCAAGGAAATTTGGGAG
rs467289	ACGTTGGATGGGCCCTCTTGGCTTGTCTTT	ACGTTGGATGAGGCAGTGTGCCCTCTCATC

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs462644	ACGTTGGATGATGATGTGGGTGAGCCCTTG	ACGTTGGATGTAACACTCAGCACGCCACCAG
rs458752	ACGTTGGATGCACCCACATCATGTGCGCTT	ACGTTGGATGCCCTTCTCTACCCAGCACTT
rs708320	ACGTTGGATGAACACAGCTGGCTAACATG	ACGTTGGATGCACAGGTGCCTGCTATCATAC
rs457954	ACGTTGGATGAACACAGCCTTGACTGATGG	ACGTTGGATGCCTCATACAAGTAGCCAAAGG
rs2411810	ACGTTGGATGGCTTAACACAGACCTTGACTG	ACGTTGGATGAGTGTAAAGGATATCCACGGC
rs3084687	ACGTTGGATGATCCCTTGAGCCAGAGATTCT	ACGTTGGATGATGCTGCTGCACACCAAG
rs69638	ACGTTGGATGTGCTCATTTGCTGTCTCTCATC	ACGTTGGATGAGAAGAAAGGTGTGCAGTGG
rs455452	ACGTTGGATGAGTGTGATGATGAGCCTGCTGG	ACGTTGGATGTCAAGTTCCCTCTCTGTGTC
rs464850	ACGTTGGATGTCTCTCTGCTGCCAGACCA	ACGTTGGATGTGGGCTGAGATTCTGTGGG
rs431472	ACGTTGGATGAACCAAGTGTGGGTGTGAAGC	ACGTTGGATGAGAGACTGCATCAGGCAGGA
rs2411809	ACGTTGGATGAGCGCATAAAGTGACCACCAG	ACGTTGGATGGCACTCACAGGGCATTGATG
rs2457094	ACGTTGGATGTTACTGTACCTTGGGTCTC	ACGTTGGATGGGAAGTCTGTATAGACGCAG
rs2457095	ACGTTGGATGTTATCAAGGCCCTGCCAGTG	ACGTTGGATGACTCCTGACCTCAGGCAATC
rs2261740	ACGTTGGATGATCGTGCCACTGCACTCCAG	ACGTTGGATGTCTCTTTTGGTAGCCCCC
rs1109180	ACGTTGGATGCCAGGCTTGATTGCACATC	ACGTTGGATGAGAATGCGTGTGCATGTGGG
rs1109179	ACGTTGGATGTGTAATGGTATGCAGACCCC	ACGTTGGATGGAGTCCCGTATTTGTCTTC
rs1109178	ACGTTGGATGGCAACAACAGCAACAG	ACGTTGGATGAAGTGTGGATTTGTGCAGAC
rs456909	ACGTTGGATGAGCTGCTTCACTGTAAAG	ACGTTGGATGGGCACTTACCGATCTACTC
rs469124	ACGTTGGATGACTTGGACACACATAGGCTG	ACGTTGGATGTGAATGCTCAGGGTGTGTG
rs469090	ACGTTGGATGTGAATGCTCAGGGTGTGTG	ACGTTGGATGAGCACTTGGACACCATAGG
rs467017	ACGTTGGATGGTCTAGCTGCCACTAAACAG	ACGTTGGATGATGTGCCAAGAGCCTTTGAG
rs469290	ACGTTGGATGTGCCCTTTGTGTGCTCAGAG	ACGTTGGATGTCCCTCTGTGCTGTGTTGG
rs469090	ACGTTGGATGACTTGTCTCAGGTGCTTGG	ACGTTGGATGGATGGTGTAGTCTCTGCTGGTTC
rs469568	ACGTTGGATGAGCACCTCTGGCTTTCTATTG	ACGTTGGATGATTCACCAGGAAATCCCAAC
rs468386	ACGTTGGATGTAATCCCAAGCCCTTTGGAAG	ACGTTGGATGTATGGAGACAGGGTTTACC
rs469349	ACGTTGGATGTAGAGACAGAGTCTCACTC	ACGTTGGATGTTGATCCAGGAGTTCAAGG
rs469099	ACGTTGGATGTTGGAAGCTGCTCTAGTTCTC	ACGTTGGATGTGAAAACCGGAGCTCAGCTC
rs456868	ACGTTGGATGACAGAGCAGGGAGCTGCGGT	ACGTTGGATGATTCACCCCCAGCTACTGTG
rs465389	ACGTTGGATGAGGCTTTGTAGACAGCTCCC	ACGTTGGATGTGCCAGTGGCTGTGATGTGGG
rs463892	ACGTTGGATGAGGCTTTGTAGACAGCTCCC	ACGTTGGATGTGCCAGTGTCTGAGTATGC
rs468548	ACGTTGGATGACTGGAAGGGGAACATGCAAG	ACGTTGGATGCCTGGATGCCCTTTATAGAC
rs654612	ACGTTGGATGACTGGAAGGGGAACATGCAAG	ACGTTGGATGTGGATGCCCTTTCTAGACAC
rs468542	ACGTTGGATGGCCCTCAATTTCTCTCTCAC	ACGTTGGATGTGCTAGAAAGGGCATTACAG
rs469262	ACGTTGGATGTTCTGAGCTGAACGAAGCAG	ACGTTGGATGGGTGAGGATCCTTTGATGC
rs708323	ACGTTGGATGCACATACTATACAGGTCAAC	ACGTTGGATGGAGGAGGAAGATGTTGTGAA
rs469089	ACGTTGGATGTTTGGAACTACCACTCAGC	ACGTTGGATGAATGGAAGGAAGGATCAGCC
rs469396	ACGTTGGATGAGTGAATCTCAATGAGGGAAC	ACGTTGGATGCTCTCACACCACTGACTCTTC
rs468723	ACGTTGGATGTGTGGATCTTGTCTGTTGGG	ACGTTGGATGTATTGGCATCGCGTATCAGG
rs467604	ACGTTGGATGACTCCTGCCATTAAACTCTC	ACGTTGGATGCTTGGCTTAACCTACAAGGG
rs338874	ACGTTGGATGACCCCAACACAGCCACTGGG	ACGTTGGATGAAGGGGCTTGGCCCAACCCAA
rs338875	ACGTTGGATGTGCTGTCTTGTCTCGCTGTG	ACGTTGGATGACACTGGATATGTCAAGGTC
rs1385803	ACGTTGGATGTCAACCACTTCCAGAAGTG	ACGTTGGATGACCTTCTCTATTGCTGTGGC
rs1385804	ACGTTGGATGTCAACCACTTCCAGAAGTG	ACGTTGGATGACCTTCTCTATTGCTGTGGC
rs338876	ACGTTGGATGTTAGGGCTGGGTGGAGGAAG	ACGTTGGATGCCAATCCCACTGACAGAG
rs189803	ACGTTGGATGCCTCCAGTTTCTCTCTCTG	ACGTTGGATGATCCTGGATTAGCCAGATGG
rs452215	ACGTTGGATGATGCTCTATCTTCCACCCC	ACGTTGGATGAGCGAGACTCCGCTCATAAA
rs641170	ACGTTGGATGATAGCTCTATTCTTCCACCC	ACGTTGGATGAGCGAGACTCCGCTCATAAA

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs584398	ACGTTGGATGTTCTCTGTGAGCTATAGAAAC	ACGTTGGATGCGAGACTCCGCTCTCAAAAAA
rs385330	ACGTTGGATGTTGCCCAACTATTGTCTGT	ACGTTGGATGGGTTTCCAGACAGTGTGTTG
rs428538	ACGTTGGATGTATTATCTGCAGACACCTGG	ACGTTGGATGATCTCATTCGCCACCTCTTC
rs371229	ACGTTGGATGTATTATCTGCAGACACCTGG	ACGTTGGATGATCTCATTCGCCACCTCTTC
rs460874	ACGTTGGATGGTCTGCGGCTAAAAATTC	ACGTTGGATGGGCGAGGCTCAACTAGAAAAAC
rs646121	ACGTTGGATGGGCGAGGTCAACTAGAAAAAC	ACGTTGGATGGGCTCGCGGCTAAAAATTC
rs468262	ACGTTGGATGGCCAGGTTTCGAAAGTTAGG	ACGTTGGATGTGGGTTGGTCATGCGGTAAAC
rs467863	ACGTTGGATGTTTCGAAACCTGGCTGATGG	ACGTTGGATGTGCCACTGTCAGAAAGACAAG
rs191434	ACGTTGGATGCCAGCTGAAACACTAGACAG	ACGTTGGATGAGCTGAAGAGGCTTTCTCC
rs2054782	ACGTTGGATGAAAAAAGCAGGCCTCAGACC	ACGTTGGATGCTGACTCTCATCTGCAGAC
rs468499	ACGTTGGATGCTCCAGGAGGGACACTACGT	ACGTTGGATGTGGCCAGCTTCTCTCGATG
rs180287	ACGTTGGATGTTGTCTGCAGAAATTACCTAT	ACGTTGGATGAAAAAGAAAAAAATCAG
rs338877	ACGTTGGATGCGTGGATGAAATTTACATT	ACGTTGGATGTTCTTTGGATCAATGTGCC
rs650665	ACGTTGGATGCCCATCTTACTCTATGATCTC	ACGTTGGATGAAAGTGCTGGGATTATAGGC
rs193419	ACGTTGGATGGCAATCCAAAGACACAGGG	ACGTTGGATGATGTTTTCATCCCCAGTG
rs180288	ACGTTGGATGTGTGACCTGGTAGCTTAGAG	ACGTTGGATGTTGTAGGAGGTCAGAAAGAGG
rs186834	ACGTTGGATGTAAAGCTACCAAGTACACAC	ACGTTGGATGAGTTGATAGGAGGTCCAGG
rs189266	ACGTTGGATGTAAGCTACCAAGTACACAC	ACGTTGGATGATGTTGATAGGAGTCCAGGC
rs189267	ACGTTGGATGCCCTCATTTGTCCTGTTGTG	ACGTTGGATGCTCTGCCTGACTCTCCTATC
rs170937	ACGTTGGATGCCCTATCAACTGTTGATGCGC	ACGTTGGATGTTCTCTCAATTGTGCCCTGTTG
rs463263	ACGTTGGATGTACTGGACCCCTTTGCACAG	ACGTTGGATGTGCCCATGCTCATGTGTTGG
rs463262	ACGTTGGATGTGCCCATGCTCATGTGTTGG	ACGTTGGATGACCCCTTTGCACAGATGCTGT
rs460454	ACGTTGGATGAAGAAGGACCGGTGCAGAGA	ACGTTGGATGACATGAGCATGGGACGTTAC
rs460455	ACGTTGGATGACATGAGCATGGGACGTTAC	ACGTTGGATGAAGAAGGACCGGTGCAGAGA
rs460505	ACGTTGGATGACCGTGGACAGCGTCTCTGA	ACGTTGGATGTGCTCTGAGGGCAGAAACA
rs931316	ACGTTGGATGATGCACACACCCATGGTCAG	ACGTTGGATGCGGTTCACTCCAGCATTTCC
rs463431	ACGTTGGATGTACACACAGCCCATGGGGA	ACGTTGGATGTTTGAACCTCACAATGTGGG
rs461542	ACGTTGGATGTGATGAAGGCCAAGAATGCT	ACGTTGGATGTGTGCCAGAACGTGAGGTG
rs463557	ACGTTGGATGTGATGAAGGCCAAGAATGCT	ACGTTGGATGTGTGCCAGAACGTGAGGTG
rs191453	ACGTTGGATGCATCCAACAGCTCTGTCTGC	ACGTTGGATGACCATCTGTAGCGCATCAG
rs2271212	ACGTTGGATGAGCTTCCCCGGAGGCAACGA	ACGTTGGATGTGCAGGTCTCGGCCAAAGAC
rs462009	ACGTTGGATGCAGGCTTCTCTGTTGCC	ACGTTGGATGTTGGTGTCCACAGTGGTGT
rs2271211	ACGTTGGATGTCTGATCCCTGCTCTGGACG	ACGTTGGATGACTGACGCCAGGCGCCCTAG
rs396474	ACGTTGGATGTGGGAGTTGGAGATGATGAG	ACGTTGGATGTTCTCAGATCCCGATCAAG
rs428901	ACGTTGGATGTCTGATGACAGAGCGAGACTC	ACGTTGGATGGGGCTCGATAATGTAGCCAT
rs452300	ACGTTGGATGAGCACAAAGCTGAAGAGGTCT	ACGTTGGATGAGGAGAGAAAGTGACAGATC
rs670256	ACGTTGGATGTAGCTCTATTCTCCACCCC	ACGTTGGATGAGCGAGACTCCGCTCTCAAAA

TABLE 12

dbSNP rs#	Extend Primer	Term Mix
rs2278221	CAAACGCTGAGGAGAAGCC	ACT
rs1650358	AAGAGACAAAAGGCGCGGGC	ACT
rs1643818	TACAGGTGTGAACCAACCGC	ACT
rs3733916	AGGCTGTAGTGTGACAGAC	ACG

dbSNP rs#	Extend Primer	Term Mix
rs1624933	GTCTCAAACCTCCTGACCTCA	ACT
rs1624857	AGACCAGCCTGGCCAACAT	ACT
rs1624832	GGCCAACATGGTGAAACCC	ACG
rs1624829	TGGCCAACATGGTGAAACCC	ACT
rs2161171	TGGAATAAGAGCCCTGCAGTGG	ACT
rs1530499	CCCCTGCCCCAGCCACAGGAA	ACT
rs888764	ATGTTGTATTGGCTATATTGTCA	ACG
rs873987	AAAACCTAAAAGATCCACGGTA	ACG
rs4078699	GACACATGATTAAACAGCAACAAT	ACT
rs870311	AAGGGCGTGACGGCCCC	ACT
rs1643817	GAAAGGGGAGAAAAGATTATCCC	CGT
rs1643816	AGGACCAGGAGTTTCCCATTIT	ACT
rs1650355	GAATCAATGAAGAAGAGAGCTT	ACT
rs888763	GGTCAGGAGGCAGAGGGA	ACT
rs1862212	GGGGTGAAAGGGAGCAGGG	CGT
rs1110514	CAGGCCCCAGGTGAGGAA	CGT
rs3797600	CTTTGTTGGTTAACCACACCC	ACG
rs3797602	GCTGACAGCTCCGGACATG	ACT
rs3797603	TGTCATTCTCCTTGTGAACCCCTC	ACT
rs3776819	CCATTCCATTGCACCTGCATG	ACT
rs252076	CAAAGTGCTGGGATTGCAGG	ACG
rs252075	GAGCATTGACAGGCATGCCCTCT	ACT
rs252074	CTGGGTGGCTGCTGGGC	ACG
rs252068	GGAAGGGTTTAAGCAAGGAG	ACT
rs252069	TGAGCACCTACTATGGGCTAG	ACT
rs194040	ATCCCATATCTTCAAAGTGATTCA	ACG
rs252070	CCTGGGCTTCCCCTCCC	ACG
rs3797606	AGCCCTTGGCCTCTCTCC	ACT
rs171667	CGCCTTTTGTCTTATGCAAAGA	ACG
rs187539	AACCTTCAGGAAAGTCCCAT	ACT
rs3836834	TCAAAATATCAAACCTACCATGAAA	ACG
rs252071	ACCCTGAGACACAGGGACT	ACT
rs252072	GCTGGGTACACTCGCGGA	ACG
rs252073	GCCGAGAAAGTCAGGGATTCT	ACT
rs379589	CGGGAGTACTGAGCACCCAGG	CGT
rs2052472	CCCCTACTGTGACTATCTCCAC	ACT
rs2052471	GTCTCCTTTGGCTGCCAAG	ACT
rs2052470	TGCCAAGGCCCTGTCTCT	ACT
rs2052469	CGCGGGGAAGTACTCGGC	ACT
rs3797608	GTCTCCTGTTCTGAGGCC	ACT
rs3797609	GGCAGAGCGGATGGCCTG	ACG
rs3822601	GTGAGGCCCTGAGATGAGAACC	ACG
rs153131	TCCCATACTCCTGTGCTC	ACG
rs751546	CCGTTCTACAGCGGTTAAGA	ACT

dbSNP rs#	Extend Primer	Term Mix
rs2279979	GGCCACCAGACAGATGTAAG	ACT
rs252060	CGTGTTCGCGCAGAGGTGA	ACT
rs3797610	CTTCTCCCTTGGGTGATGTGTT	ACT
rs194039	CCACCGTGCCGGGACATTTTTTT	ACT
rs168773	ACTGGAGATATCACGGGAGC	CGT
rs252061	CCAGCTGGTCACAGGGCTCCC	ACG
rs187537	TCATGCTAAGTGAAATAAGCCA	ACT
rs252062	GCCACCACCGTCCACAGA	CGT
rs2431255	CTGTATATTTACCCGAATTAATA	ACT
rs3797612	GGCATTTCATCGTCAGGCAA	ACG
rs3797613	CCGCCGCGGTCTCCCA	ACG
rs614114	GAACGTCTCTCAGCTTTGCC	ACT
rs252063	CTCCGCTCTGAGCCTT	ACT
rs252064	GAAAACTAAGGCTCAGAGGAC	ACT
rs252065	TGGAAGGCGAGGCTGGAGT	ACG
rs450502	GGAAAGGCGAGGCTGGAGTT	ACT
rs439252	GCCTCCCAAAGTGCTGGGATTA	ACG
rs252066	TAGCCCTCTGGAGCCAG	ACG
rs457957	GGGCCCTCTCTAAAGCTC	ACT
rs3797614	TGGCCCTCGCTCTAATGCA	ACG
rs423552	CTCGATGTTGTAGTCATCGTC	ACG
rs398829	TGGCGTGCTCCTCTAGGA	ACG
rs416646	CTCAGCAGGTCTGATCCATC	ACT
rs187450	GGGCAGACTCCCAGGAT	ACT
rs337807	GCAGGCCACTCGGTGGAC	ACT
rs337806	CCACCCAGGGGTAGCCC	ACT
rs1396438	GGCAGGCAGGTGGCCTG	ACT
rs1396437	CGGCAGAAAGCAGCCTCAAGA	ACG
rs2411811	ACATAATTTCCAAATTCACCCC	CGT
rs2898813	GGTCTGGGTGGAGGGAT	ACT
rs189256	AGCAAGGCTCTATTTGGGGA	ACT
rs173072	CTTTGCTCACATCGTGCCAAA	ACT
rs337805	GGAGCAGGAAAATTACATGACT	ACG
rs191415	GAGTCCAACGTTGTCCACAT	ACT
rs180045	ACTTGTTTCTACAATTCTCATTC	ACG
rs189255	GACTCAGTCCCAGGTTTTCT	ACT
rs652766	AGAAAACCTGGGACTGAGTCT	ACT
rs466750	TCCTTAAATGCCTTGGTTGGCAAT	ACT
rs442406	TTCTGGCTGTTGGGTTTGAAC	ACT
rs662407	AAACATCTGAAATTAAGCACC	ACT
rs592971	AGCATTTCTTTGACTGCTCTTTCA	ACT
rs457187	GGAGTATCTGTTCTTGTGG	ACT
rs459490	TTGAACATAGGAATAACCCGC	ACT
rs459668	GTCTCTTTTGTGTTTTGGAGA	ACG

dbSNP rs#	Extend Primer	Term Mix
rs462646	ATTATTCGACGGAGATTATTTGAC	ACT
rs458272	ATTATTTTTCTGTCTGGTGTGG	ACT
rs463455	CCTCCTCATTTGTCAITCTTTTTC	ACT
rs675880	CTTTCATGACATTGACACAACCTAC	ACT
rs810617	CCACAGCCTCCGCCTCCC	ACT
rs464156	GGGTTTCCAGGTTAAATGGC	ACT
rs458083	CTCCTGCTCTGCCTATCCTT	ACT
rs467333	ATTTCTTCCCCCTCCTGCTCT	ACT
rs465381	GCCTCCACAGTTCCTTGT	ACT
rs466363	GTGATGGCTCTGCACCCAGA	ACG
rs2457099	AGCGTGTGCCAGCTCTCC	ACT
rs463901	GACACAATTGAGAGCGACTTAC	ACT
rs465621	AAGTGCTAGAAGAAATGTAGC	ACT
rs463724	CCTTGCGCCATCCCCTAG	CGT
rs465242	TGTGCCCATCCCCCCTT	ACT
rs467419	AACGTAAGTCTAATGACCGCCC	ACG
rs456135	CCCTCTCCTCTTCTGGGCA	ACT
rs464536	GCTTTGGTCTCCTGAGCC	ACG
rs461898	CACAGAGCGACTCTCTCTTGGTT	ACT
rs389558	CTGACAGTTCTCTAACTCCCA	ACG
rs466752	TTCTTTTTCTCCCTGAAGCA	ACG
rs455655	CACCCCATATGGCTCATGGG	ACT
rs463435	GGGAAGGAGGTACTTAGCAG	ACG
rs2174971	GTGCCACTCTCCAGCGGCC	ACG
rs1979979	TTGCCGCGCCCCACCTC	ACG
rs411804	GAAAATCCCTGTCAACAGTC	ACG
rs1623885	CTTGGCTGCAGCACCCCA	ACT
rs1643811	GCCCTGGTACTTTAGCTCCCT	ACG
rs434430	GTGTGCATGTGTGCTG	CGT
rs187538	TAAACGGGCCAAAAACGCTAT	ACT
rs252067	GCGCCTACGGATGTCAGG	ACT
rs459319	CATGTTGAACAGAGAAACGGTC	ACG
rs467289	TCACTGAGAAATATTTGTCTCC	ACT
rs462644	GGTGAGCCCTTGGCTGTG	ACG
rs458752	TAAAGCGCTCTTACAAATCAACA	ACT
rs708320	TGGTGAAATCCTGTCTCTACTAAA	CGT
rs457954	CACCGTTTCTTATAATGCAGCC	ACT
rs2411810	GGGGACGTTACTTCTTTTCAC	ACG
rs3084687	ATTATATATGTGTGTACACAT	ACT
rs69638	CCCATTTGGCTGTCTGGAA	ACT
rs455452	CCTCAACCCAGATGCCCTC	ACG
rs464850	ACTCCTGCCTGAGTGTCTC	ACT
rs431472	GTGAAGCGGAAGGAGACTC	ACG
rs2411809	CTGCACACCCTCTGCACAG	ACG

dbSNP rs#	Extend Primer	Term Mix
rs2457094	TGGCTGGCACCCTGCACTGC	ACT
rs2457095	TGGCTCATGCTTCTAATCCCA	ACT
rs2261740	CTGCACTCCAGCCTGGGC	ACT
rs1109180	ACATCAGTGACAGTGAATGGTA	ACG
rs1109179	TATGCAGACCCCTCCCC	ACT
rs1109178	AACAACAGCAACAGAAATGAAG	ACT
rs456909	CGATTCCACGCGTGTCTG	ACG
rs469124	CCTGGCTCCATTGGTGTGAA	ACT
rs468039	CCTTCACACCAATGGAGCCAG	ACT
rs467017	CTGCCACTAAACAGATGAGAA	ACT
rs469290	ATTTCTGGGCCCCAAAGTCCA	ACT
rs469090	CCAATTGTTCAGCCACTCCC	ACT
rs469568	TGATATTGCTTGTCTGGTCTTAG	ACT
rs468386	GGTCAAGAATTCAAGAGCAGC	ACT
rs469349	GTGCAGTGGCAGCATCCTA	ACT
rs469099	GCAGGTGGAACCGCAGAC	ACT
rs456868	GGAGCTGCGGTGACTCCC	ACT
rs465389	CCCTGGCACTCGCAGACC	ACT
rs463892	AGCTCCCCCGCACCAC	ACT
rs468548	AAGGGAACATGCAAGCAAGACTC	ACT
rs654612	TGCAAGCAAGACTCGAATGA	ACT
rs468542	TCACTCACTTGATTCTGCCATC	ACT
rs469262	CACTGTGGGAATTCCAGCAGA	ACT
rs708323	TATACAGTCAACCCATTAAAGT	ACT
rs469089	CCTCGGCCTTCCCCAGCT	ACT
rs469396	AATGAGGGAACCTGCAGTTAAGA	ACT
rs468723	CAGACCCCATGCCTTGCC	ACT
rs467604	GAGTTTCTCTCTTTCACAA	ACT
rs338874	CACAGCCACTGGGAGTAG	ACT
rs338875	TTGCTCGCGTGTGCCAGCAAT	ACG
rs1385803	AAGTGAATTCTCATGGCAGAT	ACT
rs1385804	CATTCCAGAAGTGGAAATCTCATG	ACT
rs338876	AGGAAGGTGCTCCGGCCT	ACG
rs189803	TGCTTCCCCCTTCCCCCT	CGT
rs452215	TCTATTCTTCCACCCCATCTT	ACT
rs641170	CTATTCTTCCACCCCATCT	ACT
rs584398	CTCTTATATAGCTCTATTCTCC	CGT
rs385330	AGGTGTCTGCAGATAATACATT	ACG
rs429538	CCTGGGGCACAGGACAATA	ACT
rs371229	GACAATAGTTGGGGCAAGAC	ACT
rs460874	ACAAAACATCCTTCAAAAATACA	CGT
rs646121	GTTTTTTGTTTCTGTGAAAGTGCT	CGT
rs468262	CACCCAACTACTTGTCTCC	ACG
rs467863	GCTGATGGGAGGCCAATGT	ACT

dbSNP rs#	Extend Primer	Term Mix
rs191434	GTCCAGAGATCCTGCTCACT	CGT
rs2054782	CCCCCTCCATCACCTCCC	ACG
rs468499	GTGAGCCAGCAATTCTCCTA	ACT
rs180287	CAATGATCAGAACTCAGAGGTTTT	ACT
rs338877	AGAGATAAATTCCAGTGTGAG	CGT
rs650665	AGACATCCCGGCCCGGGC	ACT
rs193419	CCAAAGACACAGGGAGTAGATTA	ACT
rs180288	GAGAAATATTCTTGTGGGCTTAAT	ACT
rs186834	CCAGGTCACACACACCTC	ACG
rs189266	CACACACTCCCTCTCACTGT	ACT
rs189267	TTCTGTGCATCTTTGACGCCATC	CGT
rs170937	GATGGCGTCAAGATGCACA	ACT
rs463263	CCCCTTTGCACAGATGCTG	ACT
rs463262	GGGGAGCAGCCAGTTCCTA	ACT
rs460454	AGAGGCTGGGGACAGAGAA	ACT
rs460455	GGTACCCACCACTCTCCTTCT	ACT
rs460505	CAGCGTCTCTGACACGGTC	ACG
rs931316	GGTCAGAGCAGACACATCCACAT	ACG
rs463431	CCCATGGGGAGCACCAAG	ACT
rs461542	TGGGAGCTCCCGGGATATTGCC	ACG
rs463557	GCTCCCGGATATTGCCCA	ACT
rs191453	CTGGGCTGGGGCCCTGC	ACT
rs2271212	CGAGGAGGAGCCTGGCAG	ACG
rs462009	CTCCTCGTTGCCGCCGGG	ACT
rs2271211	GACGTAGCTGCCACACCA	ACG
rs396474	CTGGTGCCCATCTATCCTGG	ACT
rs428901	GAGCGAGACTCCGTCTCAA	CGT
rs452300	CTGAAGAGGTCTTTCTCCTTCC	CGT
rs670256	TTCTTCCACCCCATCTTTG	ACT

Genetic Analysis

[0237] Allelotyping results from the discovery cohort are shown for cases and controls in Table 13. The allele frequency for the A2 allele is noted in the fifth and sixth columns for osteoarthritis case pools and control pools, respectively, where “AF” is allele frequency. The allele frequency for the A1 allele can be easily calculated by subtracting the A2 allele frequency from 1 (A1 AF = 1-A2 AF). For example, the SNP rs2278221 has the following case and control allele frequencies: case A1 (C) = 0.36; case A2 (T) = 0.64; control A1 (C) = 0.37; and control A2 (T) = 0.63, where the nucleotide is provided in paranthesis. Some SNPs are labeled “untyped” because of failed assays.

TABLE 13

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs2278221	210	178695460	C/T	0.64	0.63	0.770
rs1650358	3608	178698858	C/G			
rs1643818	3609	178698859	C/G			
rs3733916	4318	178699568	C/T			
rs1624933	5593	178700843	A/G	0.69	0.71	0.255
rs1624857	5629	178700879	C/T	0.79	0.81	0.574
rs1624832	5639	178700889	A/G	0.41	0.44	0.203
rs1624829	5640	178700890	C/T	0.89	0.93	0.044
rs2161171	8943	178704193	A/C			
rs1530499	17968	178713218	A/G	0.39	0.39	0.861
rs888764	19887	178715137	A/G			
rs873987	21034	178716284	A/G			
rs4078699	21085	178716335	C/T	0.56	0.54	0.374
rs870311	21596	178716846	A/G	0.51	0.50	0.590
rs1643817	23379	178718629	A/C	0.27	NA	NA
rs1643816	23432	178718682	A/C			
rs1650355	24007	178719257	A/C			
rs888763	26121	178721371	A/G	0.40	0.42	0.390
rs1862212	26273	178721523	A/T	0.55	0.54	0.753
rs1110514	26755	178722005	A/T	0.29	0.28	0.572
rs3797600	27411	178722661	C/T	0.56	0.57	0.738
rs3797602	27710	178722960	G/T	0.65	0.64	0.564
rs3797603	27842	178723092	C/T			
rs3776819	28379	178723629	C/T	0.46	0.46	0.850
rs252076	29603	178724853	C/T	0.46	0.48	0.519
rs252075	31232	178726482	C/G	0.35	0.36	0.859
rs252074	31504	178726754	A/G	0.35	0.34	0.816
rs252068	32583	178727833	C/G	0.47	0.48	0.656
rs252069	32794	178728044	A/G	0.28	0.27	0.626
rs194040	32840	178728090	C/T	0.31	0.32	0.665
rs252070	33044	178728294	C/T	0.58	0.57	0.573
rs3797606	33150	178728400	A/C	0.88	0.88	0.684
rs171667	33218	178728468	A/G	0.48	0.51	0.166
rs187539	33513	178728763	C/T	0.33	0.34	0.652
rs3836834	33959	178729209	- /ATCA AACTAC CATGAA A			
rs252071	34486	178729736	A/G	0.30	0.31	0.666
rs252072	36289	178731539	C/T	0.49	0.50	0.677
rs252073	36570	178731820	C/T			
rs379589	38247	178733497	A/T	0.59	0.63	0.096
rs2052472	38477	178733727	A/C	0.05	0.06	0.508
rs2052471	38518	178733768	C/T	0.89	0.88	0.459
rs2052470	38529	178733779	C/T	0.83	0.80	0.125
rs2052469	38667	178733917	A/G	0.83	0.80	0.172
rs3797608	39781	178735031	C/T	0.06	0.07	0.578
rs3797609	39856	178735106	C/T	0.05	0.05	0.812
rs3822601	39927	178735177	C/T	0.08	0.08	0.802
rs153131	40506	178735756	A/G	0.76	0.77	0.944
rs751546	41869	178737119	C/G	0.93	0.92	0.585
rs2279979	42452	178737702	C/T	0.93	0.92	0.436
rs252060	44788	178740038	C/T	0.81	0.82	0.760
rs3797610	46059	178741309	A/C	0.17	0.17	0.858

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs194039	46846	178742096	A/G	0.41	0.47	0.035
rs168773	47712	178742962	A/T	0.35	0.38	0.266
rs252061	48796	178744046	C/T	0.21	0.19	0.508
rs187537	49441	178744691	C/G			
rs252062	49602	178744852	A/T	0.95	0.95	0.960
rs2431255	49723	178744973	A/C	0.24	0.19	0.034
rs3797612	50050	178745300	C/T	0.38	0.43	0.036
rs3797613	50171	178745421	C/T	0.21	0.21	0.941
rs614114	50477	178745727	C/T	0.50	0.53	0.387
rs252063	50818	178746068	C/T	0.57	0.55	0.313
rs252064	50833	178746083	C/T	0.52	0.52	0.806
rs252065	50881	178746131	A/G	0.22	0.22	0.857
rs450502	50882	178746132	A/G			
rs439252	51386	178746636	C/T			
rs252066	51534	178746784	C/T	0.19	0.18	0.618
rs457957	52317	178747567	A/G	0.67	0.70	0.172
rs3797614	52368	178747618	C/T			
rs423552	52970	178748220	A/G	0.90	0.92	0.215
rs398829	53023	178748273	A/G			
rs416646	53356	178748606	A/G	0.56	0.57	0.650
rs187450	53882	178749132	G/T			
rs337807	54553	178749803	C/T	0.55	0.59	0.208
rs337806	55475	178750725	A/C	0.11	0.10	0.925
rs1396438	55530	178750780	A/G	0.56	0.54	0.494
rs1396437	55691	178750941	C/T			
rs2411811	55848	178751098	A/C			
rs2898813	55879	178751129	C/G			
rs189256	56316	178751566	A/G	0.19	0.19	0.988
rs173072	56911	178752161	A/C			
rs337805	57320	178752570	A/G	0.25	0.24	0.657
rs191415	57391	178752641	C/T			
rs180045	57437	178752687	C/T	0.51	0.47	0.211
rs189255	57478	178752728	C/G	0.15	0.12	0.273
rs652766	57500	178752750	C/T	0.57	0.61	0.213
rs466750	59111	178754361	G/T	0.35	0.33	0.493
rs442406	59333	178754583	A/G	0.57	0.59	0.420
rs662407	59715	178754965	A/G	0.31	0.27	0.102
rs592971	59804	178755054	A/G			
rs457187	59851	178755101	A/G	0.23	0.24	0.842
rs459490	59929	178755179	C/T	0.21	0.20	0.604
rs459668	60052	178755302	C/T	0.20	0.19	0.648
rs462646	60240	178755490	C/T	0.43	0.43	0.905
rs458272	60359	178755609	G/T	0.22	0.20	0.523
rs463455	60381	178755631	A/G	0.25	0.24	0.644
rs675880	60456	178755706	C/T	0.63	0.65	0.591
rs810617	60724	178755974	C/G			
rs464156	60875	178756125	C/T	0.34	0.34	0.892
rs458083	60968	178756218	A/G	0.80	0.82	0.499
rs467333	60978	178756228	C/G	0.11	0.12	0.369
rs465381	60998	178756248	C/T			
rs466363	61557	178756807	C/T	0.31	0.34	0.358
rs2457099	62091	178757341	C/T	0.44	0.44	0.956
rs463901	62645	178757895	C/T	0.43	0.45	0.395
rs465621	62943	178758193	A/C	0.62	0.63	0.534
rs463724	63131	178758381	A/T	0.09	0.08	0.523
rs465242	63145	178758395	G/T			
rs467419	63406	178758656	A/G	0.65	0.66	0.647

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs456135	63427	178758677	C/G	0.79	0.80	0.686
rs464536	63554	178758804	C/T	0.36	0.34	0.296
rs461898	63661	178758911	A/G	0.30	0.32	0.411
rs389558	64093	178759343	A/G	0.24	0.26	0.325
rs466752	64153	178759403	C/T	0.35	0.37	0.446
rs455655	64409	178759659	C/G	0.87	0.89	0.536
rs463435	64544	178759794	C/T	0.68	0.66	0.428
rs2174971	65257	178760507	C/T	0.52	0.51	0.695
rs1979979	65626	178760876	A/G	0.07	0.06	0.692
rs411804	65739	178760989	A/G	0.78	0.78	0.976
rs1623885	66392	178761642	C/T	0.82	0.80	0.492
rs1643811	66720	178761970	C/T	0.24	0.24	0.924
rs434430	69177	178764427	A/T			
rs187538	69336	178764586	G/T			
rs252067	69636	178764886	A/G	0.21	0.23	0.606
rs459319	69823	178765073	A/G	0.19	0.20	0.640
rs467289	69928	178765178	C/T	0.26	0.26	0.988
rs462644	70547	178765797	C/T	0.59	0.58	0.914
rs458752	70633	178765883	C/T	0.18	0.20	0.513
rs708320	71805	178767055	A/C			
rs457954	72181	178767431	C/G	0.71	0.73	0.327
rs2411810	72200	178767450	C/T	0.28	0.26	0.252
rs3084687	72474	178767724	-A/T	0.13	0.12	0.884
rs69638	72567	178767817	C/G	0.54	0.52	0.449
rs455452	72973	178768223	A/G	0.59	0.60	0.733
rs464850	73468	178768718	A/G	0.11	0.09	0.249
rs431472	73889	178769139	A/G	0.33	0.34	0.713
rs2411809	75730	178770980	C/T			
rs2457094	75970	178771220	A/G	0.71	0.73	0.383
rs2457095	76114	178771364	A/G	0.74	0.76	0.551
rs2261740	76342	178771592	C/T	0.35	0.36	0.702
rs1109180	76449	178771699	A/G			
rs1109179	76465	178771715	C/T			
rs1109178	76791	178772041	A/C	0.46	0.45	0.820
rs456909	78042	178773292	A/G	0.55	0.53	0.444
rs469124	80758	178776008	A/G			
rs468039	80778	178776028	C/T			
rs467017	81356	178776606	A/C	0.33	0.32	0.665
rs469290	81576	178776826	A/G	0.57	0.57	0.871
rs469090	81689	178776939	C/T	0.82	0.83	0.387
rs469568	81759	178777009	G/T	0.38	0.38	0.888
rs468386	81950	178777200	C/G			
rs469349	82562	178777812	A/C			
rs469099	83591	178778841	C/T	0.66	0.63	0.264
rs456868	83700	178778950	A/G			
rs465389	83821	178779071	C/G			
rs463892	83842	178779092	C/G			
rs468548	83923	178779173	G/T			
rs654612	83929	178779179	A/C			
rs468542	84021	178779271	C/G			
rs469262	84175	178779425	C/T	0.45	0.47	0.405
rs708323	84417	178779667	A/G	0.73	0.69	0.138
rs469089	84747	178779997	C/G			
rs469396	85746	178780996	C/G	0.38	0.37	0.817
rs468723	86129	178781379	C/T	0.37	0.38	0.754
rs467604	86335	178781585	A/G	0.34	0.32	0.504
rs338874	87315	178782565	C/G	0.43	0.44	0.879

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs338875	87648	178782898	A/G	0.48	0.50	0.289
rs1385803	87764	178783014	A/C			
rs1385804	87770	178783020	C/G			
rs338876	88221	178783471	C/T	0.39	0.39	0.889
rs189803	90474	178785724	A/C			
rs452215	91148	178786398	G/T			
rs641170	91150	178786400	G/T			
rs584398	91160	178786410	G/T			
rs385330	91733	178786983	C/T			
rs429538	91772	178787022	A/C			
rs371229	91785	178787035	C/T			
rs460874	93140	178788390	A/T	0.74	0.71	0.351
rs646121	93148	178788398	A/T	0.93	0.94	0.687
rs468262	96080	178791330	A/G			
rs467863	96157	178791407	C/G			
rs191434	96313	178791563	A/C			
rs2054782	96759	178792009	C/T	0.44	0.42	0.353
rs468499	97026	178792276	A/C			
rs180287	97320	178792570	C/G			
rs338877	97732	178792982	A/T	0.04	0.04	0.863
rs650665	98713	178793963	C/G			
rs193419	99707	178794957	A/C			
rs180288	99959	178795209	C/G			
rs186834	100009	178795259	A/G			
rs189266	100020	178795270	C/G			
rs189267	100065	178795315	A/C			
rs170937	100086	178795336	C/G			
rs463263	101270	178796520	C/G			
rs463262	101276	178796526	G/T			
rs460454	101371	178796621	C/T			
rs460455	101376	178796626	C/G			
rs460505	101439	178796689	C/T			
rs931316	101820	178797070	C/T			
rs463431	102392	178797642	C/G			
rs461542	102602	178797852	A/G			
rs463557	102604	178797854	A/C			
rs191453	102896	178798146	C/T	0.11	0.14	0.123
rs2271212	189104	178884354	C/T	0.65	0.57	0.003
rs462009	189134	178884384	C/T			
rs2271211	189205	178884455	A/G			
rs396474	Not mapped	Not mapped	A/C			
rs428901	Not mapped	Not mapped	A/T	0.64	0.72	0.015
rs452300	Not mapped	Not mapped	G/T			
rs670256	Not mapped	Not mapped	G/T			

[0238] The *ADAMTS2* proximal SNPs were also allelotyped in the replication cohorts using the methods described herein and the primers provided in Tables 11 and 12. The replication allelotyping results for replication cohort #1 and replication cohort #2 are provided in Tables 14 and 15, respectively.

TABLE 14

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs2278221	210	178695460	C/T	0.64	0.62	0.624
rs1650358	3608	178698858	C/G			
rs1643818	3609	178698859	C/G			
rs3733916	4318	178699568	C/T			
rs1624933	5593	178700843	A/G	0.65	0.69	0.322
rs1624857	5629	178700879	C/T	0.81	untyped	NA
rs1624832	5639	178700889	A/G	0.38	0.42	0.265
rs1624829	5640	178700890	C/T	0.87	untyped	NA
rs2161171	8943	178704193	A/C			
rs1530499	17968	178713218	A/G	0.39	0.40	0.765
rs888764	19887	178715137	A/G			
rs873987	21034	178716284	A/G			
rs4078699	21085	178716335	C/T	0.55	0.54	0.733
rs870311	21596	178716846	A/G	0.50	0.50	0.828
rs1643817	23379	178718629	A/C	0.27	untyped	
rs1643816	23432	178718682	A/C			
rs1650355	24007	178719257	A/C			
rs888763	26121	178721371	A/G	0.40	0.40	0.816
rs1862212	26273	178721523	A/T	0.55	0.55	0.936
rs1110514	26755	178722005	A/T	0.29	0.29	0.997
rs3797600	27411	178722661	C/T	0.57	0.58	0.604
rs3797602	27710	178722960	G/T	0.64	0.63	0.879
rs3797603	27842	178723092	C/T			
rs3776819	28379	178723629	C/T	0.47	0.46	0.889
rs252076	29603	178724853	C/T	0.46	0.49	0.410
rs252075	31232	178726482	C/G	0.35	0.37	0.572
rs252074	31504	178726754	A/G	0.35	0.35	0.914
rs252068	32583	178727833	C/G	0.48	0.48	0.853
rs252069	32794	178728044	A/G	0.29	0.28	0.765
rs194040	32840	178728090	C/T	0.31	0.33	0.450
rs252070	33044	178728294	C/T	0.57	0.58	0.609
rs3797606	33150	178728400	A/C	0.87	0.91	0.119
rs171667	33218	178728468	A/G	0.45	0.50	0.125
rs187539	33513	178728763	C/T	0.33	0.34	0.709
rs3836834	33959	178729209	- /ATCA AACTAC CATGAA A			
rs252071	34486	178729736	A/G	0.30	0.32	0.566
rs252072	36289	178731539	C/T	0.48	0.51	0.400
rs252073	36570	178731820	C/T			
rs379589	38247	178733497	A/T	0.59	0.65	0.035
rs2052472	38477	178733727	A/C	0.04	0.06	0.493
rs2052471	38518	178733768	C/T	0.87	0.88	0.697
rs2052470	38529	178733779	C/T	0.84	0.78	0.036
rs2052469	38667	178733917	A/G	0.84	0.79	0.086
rs3797608	39781	178735031	C/T	0.06	0.07	0.530
rs3797609	39856	178735106	C/T	0.04	0.05	0.841
rs3822601	39927	178735177	C/T	0.08	0.08	0.904
rs153131	40506	178735756	A/G	0.77	0.77	0.964
rs751546	41869	178737119	C/G	0.94	0.92	0.265
rs2279979	42452	178737702	C/T	0.94	0.92	0.238
rs252060	44788	178740038	C/T	0.82	0.80	0.553
rs3797610	46059	178741309	A/C	0.16	0.18	0.459

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs194039	46846	178742096	A/G	0.43	0.45	0.589
rs168773	47712	178742962	A/T	0.34	0.35	0.845
rs252061	48796	178744046	C/T	0.23	0.22	0.884
rs187537	49441	178744691	C/G			
rs252062	49602	178744852	A/T	0.98	0.96	0.310
rs2431255	49723	178744973	A/C	0.24	0.19	0.108
rs3797612	50050	178745300	C/T	0.42	0.46	0.254
rs3797613	50171	178745421	C/T	0.19	0.21	0.576
rs614114	50477	178745727	C/T	0.52	0.54	0.717
rs252063	50818	178746068	C/T	0.55	0.57	0.537
rs252064	50833	178746083	C/T	0.52	0.50	0.609
rs252065	50881	178746131	A/G	0.21	0.25	0.234
rs450502	50882	178746132	A/G			
rs439252	51386	178746636	C/T			
rs252066	51534	178746784	C/T	0.20	0.20	0.883
rs457957	52317	178747567	A/G	0.66	0.71	0.162
rs3797614	52368	178747618	C/T			
rs423552	52970	178748220	A/G	0.90	0.92	0.380
rs398829	53023	178748273	A/G			
rs416646	53356	178748606	A/G	0.58	0.59	0.915
rs187450	53882	178749132	G/T			
rs337807	54553	178749803	C/T	0.60	NA	NA
rs337806	55475	178750725	A/C	0.10	0.10	0.997
rs1396438	55530	178750780	A/G	0.52	0.57	0.188
rs1396437	55691	178750941	C/T			
rs2411811	55848	178751098	A/C			
rs2898813	55879	178751129	C/G			
rs189256	56316	178751566	A/G	0.21	0.20	0.852
rs173072	56911	178752161	A/C			
rs337805	57320	178752570	A/G	0.24	0.24	0.950
rs191415	57391	178752641	C/T			
rs180045	57437	178752687	C/T	0.47	0.46	0.918
rs189255	57478	178752728	C/G	0.14	0.13	0.764
rs652766	57500	178752750	C/T	0.59	0.61	0.570
rs466750	59111	178754361	G/T	0.38	0.37	0.606
rs442406	59333	178754583	A/G	0.56	0.57	0.882
rs662407	59715	178754965	A/G	0.32	0.27	0.134
rs592971	59804	178755054	A/G			
rs457187	59851	178755101	A/G	0.23	0.25	0.451
rs459490	59929	178755179	C/T	0.22	0.21	0.671
rs459668	60052	178755302	C/T	0.20	0.19	0.712
rs462646	60240	178755490	C/T	0.42	0.44	0.439
rs458272	60359	178755609	G/T	0.21	0.21	0.755
rs463455	60381	178755631	A/G	0.25	0.25	0.783
rs675880	60456	178755706	C/T	0.62	0.63	0.741
rs810617	60724	178755974	C/G			
rs464156	60875	178756125	C/T	0.32	0.34	0.541
rs458083	60968	178756218	A/G	0.80	0.82	0.499
rs467333	60978	178756228	C/G	0.10	0.13	0.243
rs465381	60998	178756248	C/T			
rs466363	61557	178756807	C/T	0.31	0.34	0.494
rs2457099	62091	178757341	C/T	0.45	0.45	0.997
rs463901	62645	178757895	C/T	0.46	0.46	0.852
rs465621	62943	178758193	A/C	0.64	0.63	0.853
rs463724	63131	178758381	A/T	0.09	0.08	0.737
rs465242	63145	178758395	G/T			
rs467419	63406	178758656	A/G	0.64	0.65	0.694

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs456135	63427	178758677	C/G	0.79	0.76	0.339
rs464536	63554	178758804	C/T	0.36	0.34	0.553
rs461898	63661	178758911	A/G	0.31	0.33	0.727
rs389558	64093	178759343	A/G	0.27	0.28	0.762
rs466752	64153	178759403	C/T	0.34	0.38	0.223
rs455655	64409	178759659	C/G	0.87	untyped	NA
rs463435	64544	178759794	C/T	0.65	0.65	0.973
rs2174971	65257	178760507	C/T	0.49	0.51	0.476
rs1979979	65626	178760876	A/G	0.08	0.07	0.579
rs411804	65739	178760989	A/G	0.77	0.79	0.420
rs1623885	66392	178761642	C/T	0.81	0.78	0.451
rs1643811	66720	178761970	C/T	0.26	0.25	0.715
rs434430	69177	178764427	A/T			
rs187538	69336	178764586	G/T			
rs252067	69636	178764886	A/G	0.22	0.22	0.978
rs459319	69823	178765073	A/G	0.19	0.22	0.245
rs467289	69928	178765178	C/T	0.26	0.29	0.377
rs462644	70547	178765797	C/T	0.58	0.56	0.637
rs458752	70633	178765883	C/T	0.18	0.23	0.129
rs708320	71805	178767055	A/C			
rs457954	72181	178767431	C/G	0.69	0.73	0.143
rs2411810	72200	178767450	C/T	0.28	0.23	0.083
rs3084687	72474	178767724	-A/T	0.12	0.13	0.767
rs89638	72567	178767817	C/G	0.53	0.49	0.157
rs455452	72973	178768223	A/G	0.58	0.61	0.313
rs464850	73468	178768718	A/G	0.13	0.10	0.171
rs431472	73889	178769139	A/G	0.32	0.39	0.048
rs2411809	75730	178770980	C/T			
rs2457094	75970	178771220	A/G	0.70	0.75	0.157
rs2457095	76114	178771364	A/G	0.74	0.75	0.707
rs2261740	76342	178771592	C/T	0.34	untyped	NA
rs1109180	76449	178771699	A/G			
rs1109179	76465	178771715	C/T			
rs1109178	76791	178772041	A/C	0.47	0.48	0.715
rs456909	78042	178773292	A/G	0.56	0.54	0.537
rs469124	80758	178776008	A/G			
rs468039	80778	178776028	C/T			
rs467017	81356	178776606	A/C	0.33	0.31	0.480
rs469290	81576	178776826	A/G	0.63	0.66	0.427
rs469090	81689	178776939	C/T	0.80	0.83	0.300
rs469568	81759	178777009	G/T	0.39	0.43	0.234
rs468386	81950	178777200	C/G			
rs469349	82562	178777812	A/C			
rs469099	83591	178778841	C/T	0.66	0.60	0.066
rs456868	83700	178778950	A/G			
rs465389	83821	178779071	C/G			
rs463892	83842	178779092	C/G			
rs468548	83923	178779173	G/T			
rs654612	83929	178779179	A/C			
rs468542	84021	178779271	C/G			
rs469262	84175	178779425	C/T	0.46	0.50	0.232
rs708323	84417	178779667	A/G	0.72	0.66	0.071
rs469089	84747	178779997	C/G			
rs469396	85746	178780996	C/G	0.37	0.35	0.522
rs468723	86129	178781379	C/T	0.39	0.41	0.495
rs467604	86335	178781585	A/G	0.33	0.30	0.303
rs338874	87315	178782565	C/G	0.44	0.46	0.628

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs338875	87648	178782898	A/G	0.49	0.54	0.106
rs1385803	87764	178783014	A/C			
rs1385804	87770	178783020	C/G			
rs338876	88221	178783471	C/T	0.38	0.36	0.609
rs189803	90474	178785724	A/C			
rs452215	91148	178786398	G/T			
rs641170	91150	178786400	G/T			
rs584398	91160	178786410	G/T			
rs385330	91733	178786983	C/T			
rs429538	91772	178787022	A/C			
rs371229	91785	178787035	C/T			
rs460874	93140	178788390	A/T	0.74	0.69	0.118
rs646121	93148	178788398	A/T	0.93	0.95	0.477
rs468262	96080	178791330	A/G			
rs467863	96157	178791407	C/G			
rs191434	96313	178791563	A/C			
rs2054782	96759	178792009	C/T	0.45	0.42	0.514
rs468499	97026	178792276	A/C			
rs180287	97320	178792570	C/G			
rs338877	97732	178792982	A/T	0.04	0.04	0.781
rs650665	98713	178793963	C/G			
rs193419	99707	178794957	A/C			
rs180288	99959	178795209	C/G			
rs186834	100009	178795259	A/G			
rs189266	100020	178795270	C/G			
rs189267	100065	178795315	A/C			
rs170937	100086	178795336	C/G			
rs463263	101270	178796520	C/G			
rs463262	101276	178796526	G/T			
rs460454	101371	178796621	C/T			
rs460455	101376	178796626	C/G			
rs460505	101439	178796689	C/T			
rs931316	101820	178797070	C/T			
rs463431	102392	178797642	C/G			
rs461542	102602	178797852	A/G			
rs463557	102604	178797854	A/C			
rs191453	102896	178798146	C/T	0.15	0.19	0.139
rs2271212	189104	178884354	C/T	0.64	0.58	0.072
rs462009	189134	178884384	C/T			
rs2271211	189205	178884455	A/G			
rs396474	Not mapped	Not mapped	A/C			
rs428901	Not mapped	Not mapped	A/T	0.66	untyped	NA
rs452300	Not mapped	Not mapped	G/T			
rs670256	Not mapped	Not mapped	G/T			

TABLE 15

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs2278221	210	178695460	C/T	0.64	0.64	0.837
rs1650358	3608	178698858	C/G			
rs1643818	3609	178698859	C/G			
rs3733916	4318	178699568	C/T			
rs1624933	5593	178700843	A/G	0.73	0.75	0.447
rs1624857	5629	178700879	C/T	0.78	0.81	0.289

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs1624832	5639	178700889	A/G	0.44	0.47	0.423
rs1624829	5640	178700890	C/T	0.90	0.93	0.294
rs2161171	8943	178704193	A/C			
rs1530499	17968	178713218	A/G	0.39	0.36	0.499
rs888764	19887	178715137	A/G			
rs873987	21034	178716284	A/G			
rs4078699	21085	178716335	C/T	0.57	0.54	0.316
rs870311	21596	178716846	A/G	0.52	0.50	0.579
rs1643817	23379	178718629	A/C			
rs1643816	23432	178718682	A/C			
rs1650355	24007	178719257	A/C			
rs888763	26121	178721371	A/G	0.40	0.44	0.264
rs1862212	26273	178721523	A/T	0.56	0.53	0.529
rs1110514	26755	178722005	A/T	0.30	0.27	0.381
rs3797600	27411	178722861	C/T	0.55	0.54	0.840
rs3797602	27710	178722960	G/T	0.68	0.65	0.534
rs3797603	27842	178723092	C/T			
rs3776819	28379	178723629	C/T	0.45	0.47	0.662
rs252076	29603	178724853	C/T	0.46	0.46	0.986
rs252075	31232	178726482	C/G	0.36	0.34	0.666
rs252074	31504	178726754	A/G	0.35	0.33	0.604
rs252068	32583	178727833	C/G	0.47	0.48	0.648
rs252069	32794	178728044	A/G	0.27	0.26	0.640
rs194040	32840	178728090	C/T	0.31	0.30	0.734
rs252070	33044	178728294	C/T	0.61	0.55	0.157
rs3797606	33150	178728400	A/C	0.91	0.83	0.005
rs171667	33218	178728468	A/G	0.51	0.52	0.674
rs187539	33513	178728763	C/T	0.32	0.33	0.836
rs3836834	33959	178729209	/TATCA AACTAC CATGAA A			
rs252071	34486	178729736	A/G	0.30	0.30	0.942
rs252072	36289	178731539	C/T	0.50	0.49	0.684
rs252073	36570	178731820	C/T			
rs379589	38247	178733497	A/T	0.60	0.61	0.981
rs2052472	38477	178733727	A/C	0.06	0.06	0.856
rs2052471	38518	178733768	C/T	0.91	0.86	0.079
rs2052470	38529	178733779	C/T	0.82	0.83	0.828
rs2052469	38667	178733917	A/G	0.82	0.82	0.983
rs3797608	39781	178735031	C/T	0.06	0.06	0.969
rs3797609	39856	178735106	C/T	0.05	0.05	0.879
rs3822601	39927	178735177	C/T	0.07	0.08	0.838
rs153131	40506	178735756	A/G	0.76	0.76	0.981
rs751546	41869	178737119	C/G	0.91	0.92	0.526
rs2279979	42452	178737702	C/T	0.92	0.92	0.906
rs252060	44788	178740038	C/T	0.81	0.85	0.157
rs3797610	46059	178741309	A/C	0.18	0.16	0.593
rs194039	46846	178742096	A/G	0.39	0.49	0.005
rs168773	47712	178742962	A/T	0.37	0.43	0.098
rs252061	48796	178744046	C/T	0.19	0.15	0.164
rs187537	49441	178744691	C/G			
rs252062	49602	178744852	A/T	0.93	0.95	0.290
rs2431255	49723	178744973	A/C	0.23	0.19	0.201
rs3797612	50050	178745300	C/T	0.32	0.38	0.102
rs3797613	50171	178745421	C/T	0.23	NA	

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs614114	50477	178745727	C/T	0.48	0.51	0.423
rs252063	50818	178746068	C/T	0.60	0.51	0.011
rs252064	50833	178746083	C/T	0.51	0.56	0.265
rs252065	50881	178746131	A/G	0.22	0.18	0.175
rs450502	50882	178746132	A/G			
rs439252	51386	178746636	C/T			
rs252066	51534	178746784	C/T	0.18	0.16	0.451
rs457957	52317	178747567	A/G	0.67	0.68	0.728
rs3797614	52368	178747618	C/T			
rs423552	52970	178748220	A/G	0.89	0.91	0.398
rs398629	53023	178748273	A/G			
rs416646	53356	178748606	A/G	0.54	0.55	0.643
rs187450	53882	178749132	G/T			
rs337807	54553	178749803	C/T	0.49	0.59	0.009
rs337806	55475	178750725	A/C	0.11	0.10	0.889
rs1396438	55530	178750780	A/C	0.61	0.50	0.007
rs1396437	55691	178750941	C/T			
rs2411811	55848	178751098	A/C			
rs2898813	55879	178751129	C/G			
rs189256	56316	178751566	A/G	0.17	0.17	0.923
rs173072	56911	178752161	A/C			
rs337805	57320	178752570	A/C	0.27	0.25	0.582
rs191415	57391	178752641	C/T			
rs180045	57437	178752687	C/T	0.56	0.48	0.115
rs189255	57478	178752728	C/G	0.16	0.12	0.168
rs652766	57500	178752750	C/T	0.55	0.61	0.231
rs466750	59111	178754361	G/T	0.31	0.28	0.473
rs442406	59333	178754583	A/G	0.58	0.63	0.209
rs662407	59715	178754965	A/G	0.30	0.28	0.449
rs592971	59804	178755054	A/G			
rs457187	59851	178755101	A/G	0.23	0.21	0.402
rs459490	59929	178755179	C/T	0.20	0.19	0.708
rs459668	60052	178755302	C/T	0.21	0.20	0.821
rs462646	60240	178755490	C/T	0.44	0.41	0.460
rs458272	60359	178755609	G/T	0.22	0.20	0.524
rs463455	60381	178755631	A/G	0.23	0.22	0.629
rs675880	60456	178755706	C/T	0.65	0.67	0.564
rs810617	60724	178755974	C/G			
rs464156	60875	178756125	C/T	0.37	0.34	0.439
rs458083	60968	178756218	A/G			
rs467333	60978	178756228	C/G	0.11	0.11	0.902
rs465381	60998	178756248	C/T			
rs466363	61557	178756807	C/T	0.32	0.34	0.547
rs2457099	62091	178757341	C/T	0.43	0.43	0.974
rs463901	62645	178757895	C/T	0.39	0.43	0.342
rs465621	62943	178758193	A/C	0.59	0.64	0.195
rs463724	63131	178758381	A/T	0.09	0.07	0.539
rs465242	63145	178758395	G/T			
rs467419	63406	178758656	A/G	0.66	0.67	0.752
rs456135	63427	178758677	C/G	0.79	0.85	0.029
rs464536	63554	178758804	C/T	0.36	0.32	0.332
rs461898	63661	178758911	A/G	0.28	0.31	0.423
rs389558	64093	178759343	A/G	0.20	0.23	0.311
rs466752	64153	178759403	C/T	0.36	0.35	0.781
rs455655	64409	178759659	C/G	NA	0.72	NA
rs463435	64544	178759794	C/T	0.72	0.68	0.230
rs2174971	65257	178760507	C/T	0.56	0.51	0.142

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs1979979	65626	178760876	A/G	0.05	0.05	0.993
rs411804	65739	178760989	A/G	0.80	0.77	0.343
rs1623885	66392	178761642	C/T	0.84	0.84	0.819
rs1643811	66720	178761970	C/T	0.22	0.23	0.847
rs434430	69177	178764427	A/T			
rs187538	69336	178764586	G/T			
rs252067	69636	178764886	A/G	0.21	0.24	0.369
rs459319	69823	178765073	A/G	0.18	0.15	0.353
rs467289	69928	178765178	C/T	0.27	0.22	0.179
rs462644	70547	178765797	C/T	0.60	0.61	0.609
rs458752	70633	178765883	C/T	0.18	0.15	0.271
rs708320	71805	178767055	A/C			
rs457954	72181	178767431	C/G	0.72	0.72	0.882
rs2411810	72200	178767450	C/T	0.29	0.30	0.630
rs3084687	72474	178767724	-A/T	0.13	0.11	0.509
rs69638	72567	178767817	C/G	0.54	0.57	0.440
rs455452	72973	178768223	A/G	0.60	0.58	0.499
rs464850	73468	178768718	A/G	0.10	0.09	0.839
rs431472	73889	178769139	A/G	0.35	0.27	0.025
rs2411809	75730	178770980	C/T			
rs2457094	75970	178771220	A/G	0.71	0.70	0.792
rs2457095	76114	178771364	A/G	0.75	0.76	0.602
rs2261740	76342	178771592	C/T	0.36	0.36	0.924
rs1109180	76449	178771699	A/G			
rs1109179	76465	178771715	C/T			
rs1109178	76791	178772041	A/C	0.45	0.42	0.420
rs456909	78042	178773292	A/G	0.53	0.51	0.598
rs469124	80758	178776008	A/G			
rs468039	80778	178776028	C/T			
rs467017	81356	178776606	A/C	0.34	0.35	0.762
rs469290	81576	178776826	A/G	0.49	0.44	0.223
rs469090	81689	178776939	C/T	0.83	0.84	0.883
rs469568	81759	178777009	G/T	0.36	0.30	0.115
rs468386	81950	178777200	C/G			
rs469349	82562	178777812	A/C			
rs469099	83591	178778841	C/T	0.65	0.67	0.560
rs456868	83700	178778950	A/G			
rs465389	83821	178779071	C/G			
rs463892	83842	178779092	C/G			
rs468548	83923	178779173	G/T			
rs654612	83929	178779179	A/C			
rs468542	84021	178779271	C/G			
rs469262	84175	178779425	C/T	0.45	0.43	0.762
rs708323	84417	178779667	A/G	0.74	0.74	0.899
rs469089	84747	178779997	C/G			
rs469396	85746	178780996	C/G	0.39	0.42	0.569
rs468723	86129	178781379	C/T	0.36	0.34	0.573
rs467604	86335	178781585	A/G	0.35	0.36	0.763
rs338874	87315	178782565	C/G	0.42	0.40	0.564
rs338875	87648	178782898	A/G	0.46	0.45	0.701
rs1385803	87764	178783014	A/C			
rs1385804	87770	178783020	C/G			
rs338876	88221	178783471	C/T	0.41	0.44	0.580
rs189803	90474	178785724	A/C			
rs452215	91148	178786398	G/T			
rs641170	91150	178786400	G/T			
rs584398	91160	178786410	G/T			

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs385330	91733	178786983	C/T			
rs429538	91772	178787022	A/C			
rs371229	91785	178787035	C/T			
rs460874	93140	178788390	A/T	0.73	0.75	0.550
rs646121	93148	178788398	A/T	0.93	0.92	0.697
rs468262	96080	178791330	A/G			
rs467863	96157	178791407	C/G			
rs191434	96313	178791563	A/C			
rs2054782	96759	178792009	C/T	0.43	0.40	0.473
rs468499	97026	178792276	A/C			
rs180287	97320	178792570	C/G			
rs338877	97732	178792982	A/T	0.04	0.04	0.928
rs650665	98713	178793963	C/G			
rs193419	99707	178794957	A/C			
rs180288	99959	178795209	C/G			
rs186834	100009	178795259	A/G			
rs189266	100020	178795270	C/G			
rs189267	100065	178795315	A/C			
rs170937	100086	178795336	C/G			
rs463263	101270	178796520	C/G			
rs463262	101276	178796526	G/T			
rs460454	101371	178796621	C/T			
rs460455	101376	178796626	C/G			
rs460505	101439	178796689	C/T			
rs931316	101820	178797070	C/T			
rs463431	102392	178797642	C/G			
rs461542	102602	178797852	A/G			
rs463557	102604	178797854	A/C			
rs191453	102896	178798146	C/T	0.06	0.06	0.929
rs2271212	189104	178884354	C/T	0.66	0.56	0.012
rs462009	189134	178884384	C/T			
rs2271211	189205	178884455	A/G			
rs396474	Not mapped	Not mapped	A/C			
rs428901	Not mapped	Not mapped	A/T	0.61	0.72	0.002
rs452300	Not mapped	Not mapped	G/T			
rs670256	Not mapped	Not mapped	G/T			

[0239] Allelotyping results were considered particularly significant with a calculated p-value of less than or equal to 0.05 for allelotyping results. These values are indicated in bold. The allelotyping p-values were plotted in Figure 1 for the discovery cohort. The position of each SNP on the chromosome is presented on the x-axis. The y-axis gives the negative logarithm (base 10) of the p-value comparing the estimated allele in the case group to that of the control group. The minor allele frequency of the control group for each SNP designated by an X or other symbol on the graphs in Figure 1 can be determined by consulting Table 13. For example, the left-most X on the left graph is at position 178695460. By proceeding down the Table from top to bottom and across the graphs from left to right the allele frequency associated with each symbol shown can be determined.

[0240] To aid the interpretation, multiple lines have been added to the graph. The broken horizontal lines are drawn at two common significance levels, 0.05 and 0.01. The vertical broken lines are drawn

every 20kb to assist in the interpretation of distances between SNPs. Two other lines are drawn to expose linear trends in the association of SNPs to the disease. The light gray line (or generally bottom-most curve) is a nonlinear smoother through the data points on the graph using a local polynomial regression method (W.S. Cleveland, E. Grosse and W.M. Shyu (1992) Local regression models. Chapter 8 of Statistical Models in S eds J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole.). The black line provides a local test for excess statistical significance to identify regions of association. This was created by use of a 10kb sliding window with 1kb step sizes. Within each window, a chi-square goodness of fit test was applied to compare the proportion of SNPs that were significant at a test wise level of 0.01, to the proportion that would be expected by chance alone (0.05 for the methods used here). Resulting p-values that were less than 10^{-8} were truncated at that value.

[0241] Finally, the exons and introns of the genes in the covered region are plotted below each graph at the appropriate chromosomal positions. The gene boundary is indicated by the broken horizontal line. The exon positions are shown as thick, unbroken bars. An arrow is placed at the 3' end of each gene to show the direction of transcription.

Example 5

Effect of ADAMTS2 Polypeptides on Biosynthesis of Type II Collagen in Patients with OA

[0242] To investigate the effect of ADAMTS2 polypeptides on Type II collagen biosynthesis and processing, human articular cartilage from OA patients undergoing joint replacement is harvested, dissected and maintained as described by Nelson *et al.* (1998) *supra*. Type II procollagen levels in osteoarthritic patients and autopsy controls is determined by radioimmunoassay (RIA) as previously described. Allelic variations (*e.g.*, rs398829) are determined for the OA patients and controls by genotyping (See Examples 1 and 2). As type II procollagen is processed by ADAMTS2, increased levels of Type II procollagen in individuals with the allelic variation associated with OA demonstrates that this variation leads to reduced procollagen processing activity and ultimately to OA.

Example 6

Effect of ADAMTS2 Polypeptides on Type II Collagen Processing Activity

[0243] To investigate the effect of ADAMTS2 polypeptide variants on ADAMTS2 collagen processing activity, recombinant polypeptides encompassing the ADAMTS2 variation of SEQ ID NO: 2 at position 733 and a wild-type ADAMTS2 polypeptide are expressed in cell lines such as chondrocytes. Since the allelic variation of ADAMTS2 at position 733 of SEQ ID: NO: 2 will prevent the conversion of the ADAMTS2 pro-enzyme to the catalytically active enzyme, processing of ADAMTS2 pro-enzyme is monitored by SDS-PAGE analysis followed by Western Blotting using antibodies to ADAMTS2 and

methods common to someone skilled in the art. Reduced levels of pro-enzyme cleavage are apparent by the increased levels of immunopositive protein of higher molecular weight than of the cleaved active protein.

Example 7

In Vitro Production of Target Polypeptides

[0244] cDNA is cloned into a pIVEX 2.3-MCS vector (Roche Biochem) using a directional cloning method. A cDNA insert is prepared using PCR with forward and reverse primers having 5' restriction site tags (in frame) and 5-6 additional nucleotides in addition to 3' gene-specific portions, the latter of which is typically about twenty to about twenty-five base pairs in length. A Sal I restriction site is introduced by the forward primer and a Sma I restriction site is introduced by the reverse primer. The ends of PCR products are cut with the corresponding restriction enzymes (*i.e.*, Sal I and Sma I) and the products are gel-purified. The pIVEX 2.3-MCS vector is linearized using the same restriction enzymes, and the fragment with the correct sized fragment is isolated by gel-purification. Purified PCR product is ligated into the linearized pIVEX 2.3-MCS vector and *E. coli* cells transformed for plasmid amplification. The newly constructed expression vector is verified by restriction mapping and used for protein production.

[0245] *E. coli* lysate is reconstituted with 0.25 ml of Reconstitution Buffer, the Reaction Mix is reconstituted with 0.8 ml of Reconstitution Buffer; the Feeding Mix is reconstituted with 10.5 ml of Reconstitution Buffer; and the Energy Mix is reconstituted with 0.6 ml of Reconstitution Buffer. 0.5 ml of the Energy Mix was added to the Feeding Mix to obtain the Feeding Solution. 0.75 ml of Reaction Mix, 50 μ l of Energy Mix, and 10 μ g of the template DNA is added to the *E. coli* lysate.

[0246] Using the reaction device (Roche Biochem), 1 ml of the Reaction Solution is loaded into the reaction compartment. The reaction device is turned upside-down and 10 ml of the Feeding Solution is loaded into the feeding compartment. All lids are closed and the reaction device is loaded into the RTSS500 instrument. The instrument is run at 30°C for 24 hours with a stir bar speed of 150 rpm. The pIVEX 2.3 MCS vector includes a nucleotide sequence that encodes six consecutive histidine amino acids on the C-terminal end of the target polypeptide for the purpose of protein purification. Target polypeptide is purified by contacting the contents of reaction device with resin modified with Ni^{2+} ions. Target polypeptide is eluted from the resin with a solution containing free Ni^{2+} ions.

Example 8

Cellular Production of Target Polypeptides

[0247] Nucleic acids are cloned into DNA plasmids having phage recombination sites and target polypeptides are expressed therefrom in a variety of host cells. Alpha phage genomic DNA contains short sequences known as attP sites, and *E. coli* genomic DNA contains unique, short sequences known as attB sites. These regions share homology, allowing for integration of phage DNA into *E. coli* via directional, site-specific recombination using the phage protein Int and the *E. coli* protein IHF. Integration produces two new att sites, L and R, which flank the inserted prophage DNA. Phage excision from *E. coli* genomic DNA can also be accomplished using these two proteins with the addition of a second phage protein, Xis. DNA vectors have been produced where the integration/excision process is modified to allow for the directional integration or excision of a target DNA fragment into a backbone vector in a rapid *in vitro* reaction (Gateway™ Technology (Invitrogen, Inc.)).

[0248] A first step is to transfer the nucleic acid insert into a shuttle vector that contains attL sites surrounding the negative selection gene, ccdB (*e.g.* pENTER vector, Invitrogen, Inc.). This transfer process is accomplished by digesting the nucleic acid from a DNA vector used for sequencing, and to ligate it into the multicloning site of the shuttle vector, which will place it between the two attL sites while removing the negative selection gene ccdB. A second method is to amplify the nucleic acid by the polymerase chain reaction (PCR) with primers containing attB sites. The amplified fragment then is integrated into the shuttle vector using Int and IHF. A third method is to utilize a topoisomerase-mediated process, in which the nucleic acid is amplified via PCR using gene-specific primers with the 5' upstream primer containing an additional CACC sequence (*e.g.*, TOPO® expression kit (Invitrogen, Inc.)). In conjunction with Topoisomerase I, the PCR amplified fragment can be cloned into the shuttle vector via the attL sites in the correct orientation.

[0249] Once the nucleic acid is transferred into the shuttle vector, it can be cloned into an expression vector having attR sites. Several vectors containing attR sites for expression of target polypeptide as a native polypeptide, N-fusion polypeptide, and C-fusion polypeptides are commercially available (*e.g.*, pDEST (Invitrogen, Inc.)), and any vector can be converted into an expression vector for receiving a nucleic acid from the shuttle vector by introducing an insert having an attR site flanked by an antibiotic resistant gene for selection using the standard methods described above. Transfer of the nucleic acid from the shuttle vector is accomplished by directional recombination using Int, IHF, and Xis (LR clonase). Then the desired sequence can be transferred to an expression vector by carrying out a one hour incubation at room temperature with Int, IHF, and Xis, a ten minute incubation at 37°C with proteinase K, transforming bacteria and allowing expression for one hour, and then plating on selective media. Generally, 90% cloning efficiency is achieved by this method. Examples of expression vectors

are pDEST 14 bacterial expression vector with att7 promoter, pDEST 15 bacterial expression vector with a T7 promoter and a N-terminal GST tag, pDEST 17 bacterial vector with a T7 promoter and a N-terminal polyhistidine affinity tag, and pDEST 12.2 mammalian expression vector with a CMV promoter and neo resistance gene. These expression vectors or others like them are transformed or transfected into cells for expression of the target polypeptide or polypeptide variants. These expression vectors are often transfected, for example, into murine-transformed adipocyte cell line 3T3-L1, (ATCC), human embryonic kidney cell line 293, and rat cardiomyocyte cell line H9C2.

[0250] Modifications may be made to the foregoing without departing from the basic aspects of the invention. Although the invention has been described in substantial detail with reference to one or more specific embodiments, those of skill in the art will recognize that changes may be made to the embodiments specifically disclosed in this application, yet these modifications and improvements are within the scope and spirit of the invention, as set forth in the claims which follow. All publications or patent documents cited in this specification are incorporated herein by reference as if each such publication or document was specifically and individually indicated to be incorporated herein by reference.

[0251] Citation of the above publications or documents is not intended as an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents. U.S. patents and other publications referenced herein are hereby incorporated by reference.

Nucleotide and Amino Acid Sequence Embodiments

[0252] Table A includes information pertaining to the incident polymorphic variant associated with osteoarthritis identified herein. Public information pertaining to the polymorphism and the genomic sequence that includes the polymorphism are indicated. The genomic sequences identified in Table A may be accessed at the http address www.ncbi.nih.gov/entrez/query.fcgi, for example, by using the publicly available SNP reference number (e.g., rs398829). The chromosome position refers to the position of the SNP within NCBI's Genome Build 34, which may be accessed at the following http address: www.ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=hum_chr.inf&query=. The "Contig Position" provided in Table A corresponds to a nucleotide position set forth in the contig sequence (see "Contig Accession No."), and designates the polymorphic site corresponding to the SNP reference number. The sequence containing the polymorphisms also may be referenced by the "Nucleotide Accession No." set forth in Table A. The "Sequence Identification" corresponds to cDNA sequence that encodes associated target polypeptides (e.g., ADAMTS2). The position of the SNP within the cDNA sequence is provided in the "Sequence Position" column of Table A. If the SNP falls within an exon, the corresponding amino acid position (and amino acid change, if applicable) is provided as well. Also, the

allelic variation at the polymorphic site and the allelic variant identified as associated with osteoarthritis is specified in Table A. All nucleotide and polypeptide sequences referenced and accessed by the parameters set forth in Table A are incorporated herein by reference.

Table A

RS ID	Chrom- osome	Chrom Position	Contig Accession No. [1]	Contig Position	Nucleotide Accession No. [2]	Sequence Position	Amino Acid Position	Locus	Locus ID	A [3]	Allelic Vari- ability	OA Assoc. Allele
rs398829	5	178748273	Hs5_77500_34:3	1729878	NM_014244	coding- nonsynon	V245I	ADAMTS2	9509	R	[G/A]	G

[1] Contig Accession Number which can be found in the NCBI Database:
http address: www.ncbi.nih.gov/entrez/query.fcgi

[2] Sequence Identification or Nucleotide Accession Number which can be found in the NCBI Database:
http address: www.ncbi.nih.gov/entrez/query.fcgi

[3] "A" column is the sequence orientation ("F" is forward, "R" is reverse).

[0253] The following is a genomic nucleotide sequence for an *ADAMTS2* region. The following nucleotide representations are used throughout: "A" or "a" is adenosine, adenine, or adenylic acid; "C" or "c" is cytidine, cytosine, or cytidylic acid; "G" or "g" is guanosine, guanine, or guanylic acid; "T" or "t" is thymidine, thymine, or thymidylic acid; and "I" or "i" is inosine, hypoxanthine, or inosinic acid. Exons are indicated in italicized lower case type, introns are depicted in normal text lower case type, and polymorphic sites are depicted in bold upper case type. SNPs are designated by the following convention: "R" represents A or G; "M" represents A or C; "W" represents A or T; "Y" represents C or T; "S" represents C or G; "K" represents G or T; "V" represents A, C or G; "H" represents A, C, or T; "D" represents A, G, or T; "B" represents C, G, or T; and "N" represents A, G, C, or T.

ADAMTS2 Genomic Sequence (SEQ ID NO. 1)

>5:178695251-178884700

```

1      gggcctcggg ccagcaactgc ccagcgctgg gaagacagga gaccacaccc caagtggctt
61      tgacacaggg cgtgtccctc ttaaggcaca gaggagaagt gggcagccag ggctggggat
121     cctaggctgg ccctctcttg cccacccttc cccagggcca ctacacgtg gccagttcca
181     tggggcacca caaacgctga ggagaagccr tctcatgggt tcagggtgca gctcgggacc
241     ggaatggaca tgccggtgac agggatcatg cctgggagga gacaagaggc ggctccagat
301     gctgctcatag cctggccggg aaggtgaggg ctggcctaac tcaccaggcc tgcttctcct
361     caaggcccca atgccctctc taccacaggg aactgcccc ggctggcaca gctcagaaga
421     caccacagca gacagctcat agcctgtgac atctggctga cagcaggccc ccagcccgtc
481     accacacaag ccccgtaggc gttctgctg taacagccac actcgagctg gggcccctctg
541     ctttagtcaga tgtcacctgc tatggcttgt ctgacacctg acttctctgc ctgccgtttc
601     ctttcaagcc acctgacctc gccccctgcc tgctctcaet agtggccctc gagcagcacc
661     tgettccaca cctctctctg cttagggcat cctgccccct gtctcgctgg tatcacaccc
721     tccaagccca cctctctctg gggcggttct ttgggtctcc ccgtgttctc ctggggctct
781     tgagacccta gctagacctg ttccataaat gccgcagaag gctgcgaagg tgttcaagccg

```



```

841 atggggccaca ctaccctgca gaatatccca gcaggggggct ggctgggtgag gactgggagg
901 tcagagccgc cgtagaagat gagggccagt gccctacgaca gccccagggcc aggggggagg
961 tggcagaagt tccgtacgca gagaccaccac gtcaaaatgta ggaagatgga taatgagacc
1021 tatctcaact cccagctcct gtctcagcagc agggagggcgg cttggaaga tagggaagtc
1081 cctacagggg tatgggtgcc acacgagggtc tctgtctgta cccctggggag tcttctatga
1141 ggcttgccaa gcctggagcc ctgactttca gaaccccaac ctgccctcac tctgccaatg
1201 cccactagag agaccctgag gccacgcggt ccacagagaaa ggggaaggagg gccactggga
1261 ccccggggtga gctggcggtg tacgcgcctca aaaaatagaag catcgtgatc taaccttgtt
1321 cttttcccggt tgaataaaac agggaaaatct gcaattttttg tctgtgacat tttagagaca
1381 gcggtggagtc tgtgtgcctg agtggaaacgt tggcattccct cctctgccct ttgatagcca
1441 ggcttgacga tgcactgcgc tccgtggggag gctttccattc catctacttc agcatatagc
1501 gagcctgoga gatctgaaag cgcactttta tgtttttctt tgtttctctc cctcttatctt
1561 tactttgtgtc ctgtttttctc tctccggaat ctattagatc agagtgtgccc tctcgatgct
1621 agggctctgtc tctcaacttt tctctcatgt taactctttt gctcttttgt cccatgggga
1681 gatggcctgt actttttatt tcaagtttcg attttttttt tttttttttg agatggagtc
1741 tcaactctgac gtccaggctg gagtgacgtc gtgtgatctc ggctcactgc aacctccgcc
1801 tccctgggttc agtccattct actgcctcag cctcctgagt agctgggagt acaggcaccg
1861 gccaccatgc tccagtaatt ttgttatttt tactagagac ggggatttcac cctcttagcc
1921 aggatgggtct ccatctcccg accctcgtag ctgcccgctc cggcctccca ggtgctgctg
1981 attacaggcg tgaagccacg ccccgctctc acttttctgaa tttttactac ccccatctgt
2041 atccataaatt tatatacaaat gatttgttgt tctctaattt tctcataaaaa actctctaggg
2101 ttatggagtg tatcttctca aaactctctg tctctaattt tctataaaaa tttttttttt
2161 gtctctaaat tatctctatt ttctcttaga gccagtttct aaactttttg tttttgttta
2221 catcttggtc catcttttct atcgctgaag gatttcttgt gttttggttt ttttaaaaaa
2281 tctggaaccc ctctggctgc ctctcatttt tctttttttt tttttttttt tttagagacat
2341 agtcttgctc tgttgcccag gttggagtgcc agtgggacaa tctcggctca ctgacacctc
2401 tgcctccggg gttctagtga ttttctgccc gagtccctctc gatgtagctgc gatctacagg
2461 acctgccact acctcaagta atttttgttg ttattagtag gatgttgccc aggtctggtca
2521 cgaactcctg acctcaagtg atccgcccgc cttggcctcc caaagtgtct ggatctacagg
2581 agtgagccac cacaccacgc tctcttaggg gatagctctg ctccccatla gtgctccccc
2641 taagaactgt gtcggtttcc cctcttaggg acatggctcg gggcaatcagg tctacttaata
2701 caagacaana cagccaacag cactctcact cctctacccc agcatgggta gttttagctc ttttaacctc
2761 tgtttgaggg gagctatcca cctctacccc agcatgggta gttttagctc ttttaacctc
2821 gctatgctat taggtgctgc ttgatatttc attgtgtttt taaatttccc tttttttttt
2881 ttctgtgaag acaggccttc gctctgccc acaggtggtg gtgactcatag gggactacag
2941 ctcaactctc gggtccaac actcctccca ctcaaggtg cctcaggttc ttttccaagt tctgtgagg
3001 gcatccacca ccatgccagc cctcatttgt ttcatatttt tttcaagagg tgataaagag
3061 ttgagtgctc ttcatatgtt ttattgtcca ttcataattt cccccagtggt ttgagtggtc
3121 cctctgccca tcatctgact ggaattgtct ttatttctct cctccagtggt ttgagtggtc
3181 ttcaataact caagataaaa atctcttttt ataaaatttt cccccagtggt ttgagtggtc
3241 cttataatga caaacttttg aaatgtgcta ctatttatca gtttttttca ttgagtggtc
3301 ggttttgtttg gttttttttg tctctttttt tttagagaca gagtctcgca ctgctcccca
3361 ggctggagca aagtgggtgc atctcgactc attgcaact cgcctcccgg ggttcaacac
3421 attctcctgc ctacagctcc cgggtagctc ggactacagg cgcctcccac catgcccctg
3481 taattttttt gattttttat tagagacggg gtttccact gtttcacaa gatctctcga
3541 tctcctcagc tctgtatcta cctgctcccg actcccaaa tctgtggatt acaggtgtga
3601 accacggssc cggccctttt tctctttttt aaagaaaaca cctcttaact caagaacatg
3661 gagccatcca cctacgttat cttctttcat tattttgtt tctctattta gacctaaaa
3721 ccatgtgaga gctctctgtc ctatagagtg agtgagggtg acaggctctc ttttttcca
3781 aattggccccc acaccatttg ttaggctgtt tctctctctc tctctgaatt gctacttttg
3841 tgacacattta gctccctat acacataggt ccaattcagg acacrtgatt cctctctaat
3901 gatgtatttg tccatcagtg cactgcccca caacctcacc caactaaggt ggccacggcc
3961 tccctgcgcc tgacactctt cctgggctgt ggccaaagct cccggggccc tctctagggc
4021 caggggctca tttccagctt cagcctctg tggccctgca tgggtgaggg tcaacttgca
4081 tggcgggaag cccaaagtcc tgcgtgtga gaaagatggc ggaagatggc gatctgtatt ttatcatgct
4141 ggcccgctga tggctctctg tggaggtagg cccagcgcca gacattctcc aggctctgag
4201 aggggtctcc gatctcgatg aggtcactgg actgcagggg gatcgagaga aatggaagag
4261 agaggttggg cctgggtggc ccagtggggg gcccagcagg ctgtgatgtt gacagacYcc
4321 atccactggt aactgtttcc aggtggtaga aaggccctgc cgggtcactc tggagacctg
4381 ctcccactgg gctccagctc ttgacagggg agaaaagttt ccccagctac aactcactgc
4441 tgaagttctc acacaagaat ccgcagtttt tcaagtgaga tgcagaagcc agctcactacg
4501 tcatctaaag cctcgcatag tcccacgctc taacccattt ccagaagtcc ctgtgtttgt
4561 tctatgaagg acctgtttccg tccaaaggct ccaagaaagg caaattgcat gttggacctc
4621 ccccaaaagag ctacagattt tgggagggtc cgtggcaggg gaggcgaaaa ttccagtgtg

```

4681 gtgtgcaagg ccagaggaag cagatcacct tatacgttcc ccccggtgtga aggtactttgc
4741 agaaaccaac acaatggagg aaaagaaaca gcaactgaca tatttgggtgg gtgcaaaaagt
4801 aatcgcgatg ttgcatatga aagtatgggc aaagactttg gcatcaacct aatgaaaaaa
4861 cagccatata aagaacaagg gcatcacagc atgattccca caggaaactat gcatgggaaag
4921 ccccggggaa ccatcccatc ccattttgtg tcatgacgca cactcgacagc ctgaggtgggc
4981 ctccggagcc ctgctttgtg acactctctc tccagcatga catcagcagg cactcccaat
5041 tagactttgc gactaagaag cctgtcttgg acaggaacct gagactcaga gaagttaaga
5101 agcttgcaca ggtcatgaaa ctggaagatg ggagaattga gatcggaatc caagaaggct
5161 gaaacacaaa tctagatcca actcactgcc ctaaaacttcc tcccttaatac agacatacac
5221 aatatcattt caaaagcaag gcttctgtg cggcgaccct tccctttttc tcccaacatc
5281 ctccagagag gcattttgtt tctgtttctc accatagact ccaaacaccg aagttagttcc
5341 ctccaagaga accgagagcc cctctgagtc ctgggcatgt tcttttcccc ttaaggatt
5401 tatcttgataa taaaatagag atgtgcattc cttggaagct acgtatgcag gtgttttttc
5461 tcttggtcaca atgtatccac aaagttagag cactgtattc tcttttgttc tataataaac
5521 aaaaagttgg ccgtgcgagg tggctcacac ctgtaattcc agcactttgg gaggtctgag
5581 caggtggatc atYgaggtc aggagtttga gaccagcttg gccaacatRg tgaaccctYR
5641 tctctactaa aaatgcaaaa cttagtctgg cttggtggcg taactccagc aaactgggag
5701 gctgaggaag gagaatcgct tgaaccgggg aggtggaggt tgcagtgcag tbgatcacg
5761 ccactgagct ccagcctagg cgacagagtg agactccgbc tcaaaaacaaa acaaaaacaaa
5821 caacaaaggt tattgttagc aagccagatc attccagaag aagggagcac aat tcaaggct
5881 taggttaagc aagtcgttcc caggggtccc agtcaatgac tgacaagatg gcagagcaat
5941 tcccaagctc caagcacaga gtaattgtta agactttat ttcctgtccc tggagatgaa
6001 gataactcat tcaacaagg tttccatgag tcccgatgaa ataatgcata aatcagagtg
6061 tgyatgctct tgaactaagc ttgtgggaca ttgtgggaga cccagaggaag
6121 ctggaggggg gcaagggggg gtgcacttcc tgttcccatc atgcccagga agagatctgt
6181 actagctctg gccaatgagc ttgtgagcaaa agtgggggat gctctttttc ggt tggagca
6241 tctaatttcc catgtgagcc ctccagagag ctctcctttc tggcacgaga tctgycaca
6301 taaaagacag ggtgtgtctg tcaactagga cctgagtgac gatgagacc gaa tctccga
6361 ccaacacaca ttgaactata agcgtgagca ggaaactctg gtcgttttaa gccactggga
6421 tgcaaaagtgt ttgtttttct ttgagataac ctactactat ataatgtata cccagggaaa
6481 ggaagtggta tctgcatcac atccagacct gaccctctga gctctctctg agtccctgcc
6541 aggtgagtc cgctctctgg tagctctgag aggtcacgt ctgaaaggag gaatcgagcc
6601 aagaatgttc cctctccact gctctccagc aggggttga gctgggatgg cacaagatg
6661 aagggaactc cggggtggat gtgattcggt gctagagcgt gggccttgtg gttggttgt
6721 ctgcgtgaac ttaagtgg caqcccagat atgga tggagc ctgggggacc ttatcgtaaa
6781 tgaatatagc caggtcaga aagacaaaata ctgacagctc tcaactatc tggaaagcta
6841 aaacagttga tgtcacagaa atagagagta gtctgggggg agcgggttgg taagcgttgg
6901 ggaattgccaa ggctggttaa catatccaaa agtggggcaa gatgggggaa taagcgttag
6961 tgccctatag tcccgaagag cgcttggat ccgaagaaat ccttatttga tttcaata
7021 gctagaagag ttgaccttga atgttcccaa cacaagaaa tgataatgtt tgaggtgaca
7081 gatagtctaa ttaccctgat gtaattcata cacattgcat aacgtgtgtc aatatctaca
7141 cccaccccca gaaatattga caattattac atgcacata aaaaataacg taagagcaaa
7201 atcaaaaacc aaacaagaat gactttttaa catacatttt ttaaaaattg ccaactccaa
7261 tccaaatcag aggttcagca caaaggctac taataagcgc ttttctagtg gtgtgtggag
7321 gggccctctg ccagttgtgag tgtctccatc caggacaccc agcacaccc ggttagcttg
7381 ctgggctcca gggaaaaggg ccttgccttc tccagggac gtcctccagc atagtgtca
7441 agggagttcc ttccaagatg aactcaacct ccttcagggt cctaagcccc atgttagtgg
7501 accgttccct ccgatgccct ggtgggttgc ggtgacttct gctgtccctt atgttccctt
7561 cgtgaacaact cccctccttt aatctgagtc agacccagg ctccagggtc aggcagacct
7621 ggccaatccc agccaccaagc ctctcctggc accatgaagg atcagggtga acatacaacc
7681 aaagcgcccc aaggagtagg caaggtgtga ggtggggggc ctgaggtgga gcagcccccc
7741 ctgcactaca ggagaccccc caccagagat taaggaggga tgagacactc gtgcagttca
7801 gactggggct cgagtctcat cgaaggctgt taaccgagcg ggtgcagtc ctaagctctat
7861 tcaagttgttc accgcgggcc ccagctgagc tgttgcaaac cataccgat gggccctccc
7921 tggggccgct acccatgtca gctttacccc acacggtcgc cctgcttga gggaccacaa
7981 agacacaagc agaaggcacc tcccctacag aagccacagt ctgcttttga tagaattcta
8041 gtcccgccta ccgcagggca tggacattaa gagactgggt cactggcgcc gatcacgaag
8101 gctgctgcgc actgacgggt gtcatttatg cctgaagaa tctctctctc tgttgaagcc
8161 aattctgagt ttgtcctttg tcaattgcag cgaagagaa ttccagcaag gctctctcc
8221 ttttggggag cagcaggacc ccagtcaggga cactcgtctg cccatgaaga atggcgctgc
8281 atctgggggc agtgatatg tcttgggggc agttttatgt tctctcagtg gggaaagggc
8341 cctggccaa ccttttagga ggagtcagga gtccagccag aaagacatgc cccaattata
8401 ggacacagag caggaatgag agtagtgtgt aggcacagga tccggggaga gccaggctga
8461 ggagagctgg ggtgtgtcag ggaggtccgc tatcctagtg tbtgggcttg tggggagtag

8521	cgggagga	ctagaaacc	aagcgctac	aggagcaag	cagggtgcaa	gtcaatgagg
8581	caggctggt	gcacgaaag	taagtcac	ctgtgtttc	ccagttttga	aggagaaagg
8641	gtacagagga	atatttccct	attagaatag	agtgggggtc	gggtgataagg	agtggcgagg
8701	actctggcaa	accacaggtg	ccaggcacag	ctgaaggag	caacatccgc	ccagtgtcgg
8761	caggaaacaga	ggccacagagc	tgctggatca	gatttctcag	gtcaatcgtg	aaatctggat
8821	tctacagaca	ctacctaata	agtaaacact	ggctcaaat	gaaatcagtt	atgcagacaa
8881	aaacacgcgt	ctctcccgctc	accactttat	tccagaaacc	tggaaataaga	gcctctgcagt
8941	ggkCagcccc	ttctccacgt	ctcgagact	ggatccactc	tggccgatca	ctctctctct
9001	cccagatgct	ggcttctgtg	gttcagcctg	ggttcccaag	ccgtctctagc	agccaaagtgt
9061	atcttctgctg	tgcggcaaga	caatttatgt	aaatatgat	ttgtgttttt	gtcttgcct
9121	cgctttgttc	tgcggccagt	ccaagaaggc	agcctgccac	tttcgttctc	caaccccaag
9181	gaggttctgct	tgcagggctc	agagctaaag	gggtctatga	ccccctctg	gctttggggt
9241	tctgaggaga	ctccctgagg	aggggggcaga	ggggagggtc	gggggtcgatg	ggcagaggga
9301	ggggaaagctg	cactggggcc	aggggagccc	tgggaggcag	gacttcaggga	gacatgagag
9361	tgtgctagac	cttgggagcc	tgggtcccg	cagggggtac	cccaagaagt	agaggagagg
9421	gttacctcag	tgtccgggca	ccccaggact	caacctgcc	ctgggagctt	aactgacaag
9481	ccacagagct	gcttgtgctt	ctctctctc	tctgccaacg	tcccaacatc	tccaagtttc
9541	cagccctatct	ctgacgggca	cagaggagcc	caggggcgga	gatcttctgc	tcttgggtcc
9601	tggagcatct	aagggtctatg	gggtaccctat	gaggcccgga	cattgacctt	gtgcccgagg
9661	ggggaaagctg	gggtctgtgtg	catctcttaa	acagcccttt	gttgcagctg	gggaaggagt
9721	gggatagaaag	gggacagtg	ggcactgaga	tgcattcagg	tgtggctggt	ggccagctag
9781	gaggggtcagg	acctgggtggc	aggagctgag	aggcatcagg	gctgcagggc	gggagcaatg
9841	aggcagagagg	gggacacagtg	tctggctttg	gggatgtgta	gcaaggcgga	agaggccggg
9901	ggggaaagggg	acgggaaagt	ccagaggaa	agcgtgggtg	ccgcagtgca	ggaaacatg
9961	tcacgggcaat	gtctttccctg	gccagcttgc	tcgttgggga	gaagctgggt	gttttgcctt
10021	gggagctctc	ccacttctgg	gggtgcagcc	cttgggggtg	atcagtccca	agccgtgact
10081	gggacccatg	ctaggcccat	ggagcacagg	cttgaaaggga	ctgaaaaactg	gctgaaagtgg
10141	gtttctccgtg	tggcagcacc	caaaacaaac	atccgcaaac	ctcaggtctat	gcaccccgga
10201	gccccctttc	ctccaccacc	tctctccact	ccatgagagc	tacaggcgagc	cttgggttct
10261	gcatgcctct	aaggactcct	ccagaaaga	aactgatgtc	caggcagaga	catcggggca
10321	ctgtgacagg	tctcgggccc	aggagggaag	tctgggtcgg	ccgtggacag	gggaagggcc
10381	tcgtactcagg	aggcagtcgg	aaccacaagg	aagagccaa	ctcaggtctc	tcctcgtaga
10441	ctccacagcca	aggatactgt	ctcgacatgc	agtgaaactc	ccgtgtgggtc	acttgggctc
10501	cttcggggacc	acctgctggg	atgcgacctc	cttgtctctc	ctgcacacctg	gectgcctgc
10561	aggaggccag	atgggtccct	gctgcctggg	gcacccacagc	ccccacacatc	ctgtgcccctt
10621	gctatgtatc	aagcctggga	actgtgctc	ccctgagcga	gaagaagaca	catctctcat
10681	atctggatct	ctccagaag	tgaggggaggc	tgcactgctt	gtttccactg	aggagttctt
10741	caccagctcgt	ccctggctcat	ctacgtccct	ccagctcctg	aatgcactctg	acttccagct
10801	gctctgctcgg	ctctgcacgg	ctcctctagg	aaagtctctc	tgtgtgactc	caacctgtcc
10861	tagctgtgct	gagggggggga	ccaggagtga	gccttagcac	ctcgggctct	gcmaaaggctc
10921	tggctcgagg	accagccagg	gactttacag	ggcagccagt	ccctcccgag	gggacggcgg
10981	ggcaaccacag	ctccagagca	gggggcggta	gatgggtcgt	atagaccggg	gtatgaaact
11041	tgaaggccct	ggggctgcat	ggtagctgca	gggtccacct	cttgagtgag	gcaggggcctc
11101	tgaaggcgaga	ctcttcaacag	cgccagcaga	tccagaaact	cccaaggtcca	ctctaggccg
11161	ccagaagaagc	caggcgaggag	aagcctgcag	tagggctgtg	gatgaagtcc	agggggcccat
11221	ccagcccaagg	atgcaggagg	agtccctgag	tagctgaccc	tagctcagag	cagggggaaca
11281	gatgagcttc	caagaggctg	atcggaact	ctacaggggc	ccaggagacag	tcgtcaccctc
11341	tcatgtgaggg	ctctggccacc	ctctgttggga	gggtgacatg	accactcgag	ctctggggctc
11401	ccaaactgggc	tcactggccc	gagtgctctg	ggggccctgga	ccaggctggc	tcttccagag
11461	tgcagggtct	tccttggcca	tcatgaagtc	ccacctatgt	ccactatggg	acctcctctg
11521	ctatctgggt	tcacagagct	gccagatgc	aggaggcaga	cactcgtcca	ggccctctggg
11581	ccctcccccac	agcagcacgg	ggagcctgt	cccaagaagg	atggtgttct	ctccgggtct
11641	agcctgttctg	gcttccctgg	ccggggctg	aaagggtctg	aggacatgcc	tggaaagacc
11701	agagctccca	ccatgaggtg	gagaggcgag	atctgagggg	cttagggagg	caagagcttg
11761	atggctcccca	cccttgggtca	gaaggtcact	tgtctgaatg	ctccccaggg	ctgtcaggag
11821	gttctggccag	ggcaagggtg	aggaaacact	gcggggccctc	gaacctcagg	cccagcctgc
11881	aggggtctcgg	gtaacccccc	gggctcagg	gaacagccag	gagcttaggg	ggggcgtgagg
11941	ctcttgggtgg	acgagtggc	tgcgtccagc	tccaagtccc	agccccaact	ctctggcgggg
12001	accttggtctc	ttagctgccc	gcggggctccc	tcctctctgg	tgcgatcggt	caactcgccag
12061	atggggcacc	tactggggat	gtcctgtgtg	ctgtctgctg	ctcttgagtc	actaaagcct
12121	aagctcctctg	agggccgggtg	gggtgtcggg	cccgagctctg	tggcccaagc	ctggggccagc
12181	cttgcctctc	gtagcccccac	tctccctctc	acggagcaat	cccggtctct	gctctccatg
12241	ccaggaatgag	gcagcgaaag	cggccagagg	accccggtctg	ctggggcctc	agtgatgaaag
12301	gaatgaacac	gcagacccctc	aggagccagg	ctgggagggg	aggagagcgc	cgggtcgggg

12361 gacgttcttcg ggaggcctgg ggaagtgaac ccaaaaggcc gccccggcag atgctgacc
12421 cagcaggaag tcccgctcgc caaggagcgc ccgcgtaggc tcaggcttgg cacacgacac
12481 ctcgcttcca ctccatttcc ctacccttcg gatgacgaac tgcgtctcag acaactgaaa
12541 ggactcgccct gactccaca gccggaaaga ggtcggggtg ggtcctcccg cgcgggcggg
12601 gcagaggaaag ggccggggccc gagcggccac caggtggcgc tgcctgcgcg agaacagcgg
12661 cccggggcgcc ggacgcacccc ggccgccaga ctccgcccgc gcaccggcgc tccggggcgg
12721 ccaggggccc caagacaagcc tctgggctcc accccaataa catcaattat gacctgagca
12781 aaaaattcttg tccaatcgt tgcctccoga gatttccatt gtttgcctcg gaaggagctg
12841 aatctatctat gctcttccct ccttcagtcg aggagacgtg ggaggaaccc gggatgggga
12901 tcagcagcct ggagggcgcc tccgctggaa ccgcgcgcgc gcgcggcaga ggcctgcctc
12961 gctggggcaag gcgcactcag ccagcccgcg aggcgcgcac gcgacggagg ctccaccaat
13021 ccactgcagg tggcaggtga ggtcaactgg gcagctctag gcccgccga aatctctcat
13081 ctgctctggt ctggggcaag ttacccttact tgcctccacc tgcaaaatgg gcaacaactg
13141 agctcgcagc ttggagctgg ggggcttcgc cccatgacag ttcaggcagc gtcgactcag
13201 gcgccacaaa tgcctcttgt ccttgtctct cccatgacag tgcgctgctg tgatactgat
13261 accagctccc ggggaactgg ctaccgcctg ggcctggctc tgatactgat gggatataac
13321 ccttcagcaaa gcttctcaac cgtcccccac ccagaggggg cctctgcctc ggggatgcgg
13381 tctcaacagg gtgcaggata ggacagcagg gaagtgggt tctggtagct agaccaatct
13441 tctcaacctc agcactccag atatttgggg ccagataatt cctgttggtg gaggcgggga
13501 cccgtctcat ggacagatgt tgagcacaca cctgggcact ctttgccaga accagctgga
13561 accctccccc tagtcgtgac aacccaaaagt ctctccagac actgcgcaat tgcctccggg
13621 tggggggggg ggtggcaag tcaactcttg ttggggcgaa ctgcgctcca ggggacaggg
13681 acaaggagcc tgaagctggt gctcctgacc agccgggcat acagagtcga gccctctctc
13741 ccagagcaga gtggctccac cagcgtgtgg ccccgccgag ccgcctgcgc tgcctgagcc
13801 agggctctca ctgtcaaatg gggatgatca cagcgctcac tccacagggt tgctgtgggg
13861 gtgaaggcat aagatcacat aaagggtctc agcggctgcc ttgcccagaa agacgtctg
13921 aaacagtcac ttgcacgaac gcctttatga ttgtctgcat tctgttgtg tctgtggggt
13981 tcacagacct cctcttatgg gggaaaactaa ggcctccatga tgttaaatg ctacccaag
14041 gtcccagag catgtgcaaa gccagaatga atggcaggga aaaggtccca ccatccctgc
14101 gtgtgagca catgttctac cgtgcccccc agctgagtgt agctggggt cccccaaggc
14161 tgaagctctc ttgttcagg cagggtctac agatcaaatc tcatcacgac atcgctccag
14221 tggagctggg tggtagggcg tgcacctcat ttctctgcca gaactcccat atcggtccag
14281 cattttgtt ccagccatct tccccaagg ctggggaagt attcccagag agagccaaag
14341 tgccagaaca ggaattccaa cttgagcctt cccaggcatg agccagtgca ggcctgtggg
14401 gaggcagaag aggggctctc tctagaatc tgaatgtgt gtcccacca tttgaaaggt
14461 tgaagctga gaaccaatg tatgttatga agagatggtg tctttgagag gtcattccag
14521 catgaggtg gactctccat aaaaagagat caaaaagaagc ctgaggggagc
14581 ctgtctcacc ttccatcatg cgagggcaca gctagaaggt gccctctagg aggaatgggt
14641 ccttaacaca cataagttct gctggcgtcc gtctcttga ctccccagc cccagaacag
14701 tgagtaaac aattccattg gttataaatt actcagttca aggtatttct gttatgag
14761 cccaacaga ctaaggcacc tcccaaggcc agtgtccagc cttagtgcct cctagtctc
14821 tggaccagag taggtctggg tgaggagagt gatggaggt gatgtgggt ggcctatcat
14881 cctgcacccc ctatcaggt ttggggtagg acccagatgc ctgaggagga aggggctgt
14941 cctgggggtg aggcattggg ttgtgttccc acctacccc ctgcccaga tccccagggc
15001 cctrgatata tctcccccag accttcaact cctggctctc ctgcaccat gctgggccc
15061 tctaatactc ctaccctata gttgggtggg tgggcacacc aaagcccccac ccccgaccc
15121 tctcttctcg tgcctttaca taagagactc acatgcagaa ttctcacag ctcttcaact
15181 caatgctgaa aatatccagc aagcatcca cactttgtct tgaactggt ttttataaca
15241 agacacttta tgtcctctc tgatgctccc caagtctctg atttccatt ctcttgattt
15301 tcatcatctt aggaataatt taaatgaga acagttagt agctttgga cttctccaaa
15361 aatatcaatt aaagccagt tgcattgcaa cctcaagtgt tccgagtgt cctctccctg
15421 gctggtgta tattgtataa aataagatca tctctttct tctgctttt ttaagaagt
15481 gggagaaaaa ttgttttag gaaaaactg agataaactc gaggtctctc ataaagatga
15541 agacaaaact gtcactcaaa caagtgtcac caattttatc cccactggag cctgtgagg
15601 cccgtgggac aggcaatggg ggagcggcag gtggggagct ggtctattt tgcacaaat
15661 tgggagtgaat aacaggaacc gagtatctgg gccacttct gcccaagtct gggatcgtg
15721 ctacatccc ccaagctctg tccccaccgt tggggggaca gctcagccc cctcctgact
15781 ccatcgctct ggttttgcca tgcataaatt ctctagaga cggggggctc attcttgac
15841 gggaagcctg tggttccttg aagagcagct catcagaatt catccccac attgtccaaa
15901 ctttgtcccc agagcccac cgagaatccc tggctgtgct tctctgtgag caaaaggctc
15961 ctgttcccc ttgcacagga gctaggcctt tgcctcccag atcttggctg ggaagctccc
16021 atctggctg ctccatccat ggtcccctat tgcctgcac cctcgtgat tcttggggct
16081 tgcacacac tctgtcgtg tcttgtctg ttgtactgoc ttgcacaag gctcctctcc
16141 cctcctggt catgcacct cctgggata aacagcagct gctcgggaag ccaggagctc

16201 tgggctggcc ggaatgctttg ggggcacaag gatacatctc cteccacgtt catcacagac
16261 ctccagagacc tgaaaaacagg gctcccccac aatgtaagtc tggctcctggt
16321 ggacacaaac tggctccagtc cacagtggaga taagtacaga aactagaaac tttctatagca
16381 attggagggg ttgagttttat gtccacagaga ttaataacaga aactagagttt ccttagagttt
16441 ttggataccct cgggtttttgt cacatttgtat tttccagcta attcatttat aacgtatggt
16501 accgaagatgt tggctgtgtga tgggttttga ataggaaaaa gaaaaacagg aggtctttgg
16561 catctagggtg cactgtgtac tgcgtcctcc aggatggcct cctcctggc cactggggga
16621 tggcagatgg cctcctcggg cccacagagg ctgctcccaa ggcctgtggc cccacagaca
16681 ttcgctaaagt cagccagggc ccatgaccag aacctattt cctagcggct gtgcctctca
16741 ccccaagctt tcagcccagc cctagcaggt gccactcct ctgctctact atttctcgca
16801 cttttgagct aggtccaaac taaagcaacg aggaactata gcctcggtc acttccctc
16861 tcacagccca gaacccccca gagaagggca cagcagcggg gaggcccgcc tcttctggca
16921 gttctacact agctgatcaa ctcaggagat ccccggggat gaggcgacc catgggact
16981 gggctccctt acccccacat ca ctgcctcctg tggttgatc tggggagctt taggaggaag
17041 ggggggtagg gccaggcaga aaaaaggtggg cagaagtatgg ggtgagcttt gctgagcccc
17101 actctgggca ggcattgagtt agggacagaa aaaaaggccc aagatgctgg aatgttccag
17161 aacattccag gaggccaatc atccctgaat tcccggggtc tgacagaggg accgctgtg
17221 tgttgtccag caagctaggc tggggggcct cctgggtgt cactgggtcc cctgggtccc
17281 acacccctca caacccagca ctgcctctcc cctgcctcct cagctcgggg tccactgagc
17341 cctcgggggc gggggtaagc tggttccatc tggagttagc tctggaaagc tggctccagg
17401 ttcgggttct gtcttccaat cctccaccac catcagttct caagccctt gcccagagga
17461 ggccggaagca aggcagcct cagtccactt atgctctc tttctctt tttactgggt
17521 cggcagaagt gcccgggtgc caccacaggg atgagctgcc ctggtttgct ggctatgagt
17581 ctcagcacc cccaggcctg cacatgtccc catgttgt cctggaaagag caggctgcac
17641 ctgacaaacc gtcccctgcc cactgcccc a tcccaactc tgccaaaaat ggaagtgggg
17701 ccccagctga ccccagtgag agcagagctt ctgaggggtg atgaggggtg gggctccagc
17761 caaagcagaa cttctcgagg caaggcgctg gaggggcagg gggcgaaggg gggctggcca
17821 ggggggtggg gctgtgtaac agggggcgct tgggtctcca tgcaggggag acccagttcc
17881 tatggcgagc tgggtcctca ctccaaagt tcccatctc ttcaggttga cccgctggag
17941 gtggagcccc tggccagcgg accaggaRcg tggcttagc tgcagaggtt ggaacacaga
18001 aacctcatca ctctccccc gccactgca gagatagct gccaagggaa agcatgaatt
18061 ctctttgatg tttctcatga acccagaaat agcgtgacaga gctgtgagc gctgtggg
18121 atgcgccggg agacctttct ccagagtgt gaattcagtc aaacgggcat atttggatcc
18181 taccattttt aaatgatttt ttaaaaaatc alttaaaaaa tttaaatttc acaaatttaa
18241 aatgcagtg tttgtttata aaagaaacat gctgggtccag gctcagctgc tccacctgtg
18301 aatcccgatg ctatggggagg cagagggcagg agaactattt gagcctggga gttcgaagc
18361 agctggggca acagagtgag accctgtctc tacaataaaa aaaaagaaa aaaaattagt
18421 gagtggggtg gtggtgctt gtggttggga ctacttggga gggcgagatg agaggaatccc
18481 ttgagcccaa gagatggagg ctgagtgag ttatgatcac accactccac tctagtctgg
18541 gcaacagagc tagaccccat ctcttaataa taataataat aataataata ataatcga
18601 aaacctagct catcatattg aatttaaaaa ctactagtaa gtacataaaa gaaaatgaaa
18661 ataatatgac ctactgccca cagagaanaat ctgttcacat ttgatgtgtt ccttccagtc
18721 tttttgttga atataattgt acacaaatg gagaaactga gactcagctt gactcagctt
18781 ttcaaaacac aaacagcttt agactgtaac catctccaaa tggcatataa catgtttcga
18841 aaatgcaatt ttgtggcagc cttaactct tttatcat tttatcataa aatgcataga
18901 ccaactattat tggacaatat aggtttgttg ctgtgttcca catgagatg aatggcatga
18961 ccttgtgctt gttctcaaaa ctgtggcacc actgtggatc actgctttg agcactttg
19021 tagaataatga tggctggggt cacaagaatt aatgtctggg ggcacaaact agaaagatt
19081 taaaaggcaa atgcgggtggc tgcgcacca taaccccagc aacttggggg agcaggcgg
19141 atcaactgag gtacaggagt caagaccagc ctggccaaac tgggtgaaac ccatctg
19201 taaaaataca aaaaattagc tgggtgtggt ggcaggtgcc tgaatccc gctacttgg
19261 aggcctgagg aggaataatc ctggaaccca ggtggcagag gtgagtgga gctgaggtc
19321 cgccactgca cccagcctg ggcaacagaa caagactcca tctcaaaaaa atgaagtta
19381 atgaagaaaa aaagaacatt ttaagggca gtgataataa tggctacatt cctctagaat
19441 actctgactc atactcaag tgtgctgct tcatctacc tcatcataat cctctctcat
19501 cttacagggt ttgatgcttt ataatgtgt tattattat tttcttaaca gttagagttag
19561 atatttttcc atatgggctt tttcttagcc acttggattt tttctttgt tttctccat
19621 tgacaatttt gtttcttat gatttagagg aggttttata tattaaatga tattatacac
19681 attgcaaatg ttttctcgc ttgtcatatt ctttctatt ggtttttga accaagtttg
19741 ggtattatca tttatccaat atatccat tccctttat aatccttca tattgattta
19801 tgcctagaaa gctcctgaca accaatatt tcttcttgat ttgtgtgatt taatagttg
19861 aatgttgtat tggctatata ttgtcaYat ataatttaa tttctattta ttaaatttat
19921 gttatatatt tattataac acataagca tttgtgttt actacggtag ttatttga
19981 agaggggtat tttcaactat ttatttcaac tgaactgtct atgaaagtt atgcacatat

20041 ggcactattt taattactct cactttacaa gatgtttgga ggaagtcctc tacttgcttg
 20101 ggtccacacag tctcttcggc cactcttgac cgttcatggt cctttggagg aagtcctctc
 20161 gcttgcttggt tccgcagctc gtctcagcca tctctgacca tcttgaggaa ccttgaggaa
 20221 gtctctcttcc tgcctgggtt ccgcagctgt ctagccattc ctgaccttcc tctgtctctt
 20281 ttggaggaaatg ctctctcttg ctctgggtctc gcagctgtct agccattctc tgacctgtca
 20341 tgttctcttg gaggaagtct ctctcttgct tgggttcgc agctgtctca gcaatctctc
 20401 accgttcgat tctctttgga ggaagtctct tctctgtctt ggttcgcagc ctgctctcagc
 20461 cactcttgagc tgttcatggt cctttggagg aagtcctctc ctgctctggt tccgcagctc
 20521 gtctcagcca tctctgaccg tccattgtgc ttggaggaa gtctctctct tgcctgggtt
 20581 ccgcagctgt ctcagccatg ctgaccgtgt catgttctct tggaggatgt cctctctctg
 20641 cttgggttcc gcagctgtct cagccattct tgaccgttgc tgttctctga tgggtcttag
 20701 aatcacttca taacatttcc aactatgaat ctgggaaagg ccaacctctc gattatacc
 20761 agcttcccat aggtacagca ttgtgtctt cattttgtag cactttcttc ctatcattaa
 20821 ttaggtacaa ttctcttcca atgggtccat tgtcttgta tttctccac aggtattcta
 20881 catttttggt tactaccata atggaattt tctctccctg tttctccac taccaggctt
 20941 gttgtttttg gtccacacag aagctactga ttttttaaac aatggtctga tttttggcca
 21001 catcaccaga ctctctctct aactctacaa gttRaccgt ggatttcttt aggtttctcc
 21061 gacacatgat taacagcaaa caatRagatg ttgtctctgc tttttccaa tactagttt
 21121 atgtacacata caagtatcta ggaatttccag aaccacagta cagctgttat gggtcttttt
 21181 tttgtttgtg ttctccaaagc tatctccagg cagtgctctc ccagagcagg tccagcaggc
 21241 ttcggatgag ctgagcaaga agcttccgtg gagaaagttt ccagagacaa cctgactatc
 21301 tccgactctc gagggtacac catgcacagc tccacaggaa agtctccggg aaatttacc
 21361 cccaggaatg cccattttcc ttgtttgac cagcattttc cagactctct tgactctggg
 21421 accctctcat agtctctgtg aacatctact catactctaa gaatttgggg tgcagacacg
 21481 ttttagggtaa atgtgctct tttccgcct ttaatacaaa tcttgggttg cctgtgtccc
 21541 tctgttacta gttttctggg gatctggaaa aggaggagtg cccgtgaggg ggcctaYgggg
 21601 cccgtcagcc ctgtccctgc cagagggtct acaactggag cccgtgaggcc tctctccagc
 21661 gctccccata ggcagcaagc tctcatgctc tctcttggc cctcttccg ctgactccag
 21721 acccccgagg gctccagacc cacatccaca tcttctggc ctctctccgt gggatggcac
 21781 caggtaacct taactgcaca tctcagctga cctcttcagc tctctccctg tctccccaa
 21841 tctcaggggg tgataccatc atctccagc ctgagagtct cctccgtctc accaccctag
 21901 tctccacatct tctcgaagt gctctctctg tctctctgtc tgtatccctg gaccaggcc
 21961 tggctccagc ccagctgcca gcatccccca ttgatggcca gtgactgtct cctaagccac
 22021 tggctccacc tggccctctc gcacacagga gccagtgatc tctctagaaa gttctggagc
 22081 atcgtgtcac tggctaaagc tggccctcat ggcagccaa accctctggc gggagacga
 22141 ggtctgtctc gctccttgcc atcggtctct tccgcagctc tccgcatggg gctctggtt
 22201 cctgtctctc tgtctgtccc tctgggttca gctgccctcc ctccctccc tggatgtgaa
 22261 ctacagcat ctctatgctg gtgacttcac tggaccatct cagagtggtg cctctggtt
 22321 cacttctctc cgttgtctt cctgaccttg taaggaaaga ctgctgacc cgcgtctctc
 22381 ccatccagcc ctgagccctg gtacaaatgt tctctgtct cctctgctc tctggaagc
 22441 gctccctgag gacagagact atgccagtga atattccagc cacagcactt cccacacaga
 22501 agccctctaa acagattttg tagagagtag agttttgta tttaggaact gctgtgtag
 22561 actgtatgaa tagatttctc gatattaac aattccaca tctctaggat aagccttaga
 22621 tgatcacggg agaagtttct ctattataac aactcttaat tctctgggat gctctggag
 22681 ttttactctc atttttagaa gtgagacagg tctgtacttt gtctgcagg gctactcttg
 22741 gcatcttttg gataagggt ccattgact ccatggtag catgtgcaa atgtatgga
 22801 acatgagttt aacagggtat gtaattgctc tggctgata atgtgtgcaa atgtatgga
 22861 atgcactggc aaaaacttgg aaaaaggtag gaaactgagt tccctctgag tccctctgag
 22921 ttaaggcata tcttccattt ttggagatca gctttgaaat attcacaact caggaaact
 22981 tcaattctcc caggatctcc gtgtccctgg catagagat cctcaactct cactctcagg
 23041 gtttgactct ctctgtcaaa tgttctgact gtatcacat actctctcta cctacttagc
 23101 cgagttaagg aaaaaaaag aaaaaaagaa agaaaggctc tagggtcggg actcaagtag
 23161 atagagctgt ctaaggagta attaataaag cagggtcttat tccatcaagt actcaagtag
 23221 gggaaagagc agacacttt caggccaaat aagatccaaa gctttggcag cctctgtagg
 23281 caattaaacc tagtcaatta caaaattctt ctttttaaac aatgtaggag ccaactgggg
 23341 gaaaattgcga gccgccact ggacccaga ccaatcaaKt ggataactt tctctctct
 23401 tctagcccca ggaaccaggag tttccattt tKggaaact ccagaaagag atataactc
 23461 gctctctctc catgcacaa gcttgagaa tctcaaaaag ccagagatc ctggccagga
 23521 agggacagcc tggaaagacc cgggtttggg aaggcggaaa cagggaactc cactggaggc
 23581 aggcagccag ccacatggga tagtccaaag atccccccg acctctcccg ctgtggaatg
 23641 gagagctggc accggatgac tggggtgggg cccaccccca ggcctctgat ctgttggtc
 23701 caaagccccc cagcaaacac ctgggataga gggctcagcg acctctctc tgcctctcta
 23761 gccagatcc actggcagag gatgccttgc acatccctcc tccactgatc atccctggag
 23821 ctcgaagtgc gcgggtgca gagagccttc caacagcttc atgtgggca gatggccctg

23881 catgctttcc tttctgccag gagaagatc tttttgtagt cacattacat atactctttc
23941 tgatctttta gtttgtgctt tctctctttt tttctagatt aaataggta gagggttgc
24001 ttttccMaag ctctctctct catgtattct atgtgtttct tccattttctt aaattatfga
24061 tttctgtctt tttctgaacta tttctctatt ttgtcttcc ttaagttttt ggttttgtt
24121 ctgttgaatt ctgttcaaat gtactgagtt gaaggcagg ctaattgtgt tcgaacatat
24181 caccacaagg aagaagctttt ggaaaacata cttctgaatt cctctgaata cagctttaac
24241 caccattccaa aagatttgac agtagtgctc ttgtttttat tagtvtccca aatagvtctgt
24301 actgagctcc taarattctc tttctgttgt ctgttaccct gggtgcagg cagtggtgtg
24361 atctcagctc agctcaacct ctgctctccg ggttcaagca atctgccacc ctcagctctc
24421 caaatagctg ggactacagg cacatgccac cacaccggc taatttttgtt atgttttgtg
24481 tagagacagg gtctcaccac gttagctagg ctggctctcaa acctctgacc tcaagtgtac
24541 taccaccttt ggctctccaa agtgcaggta ttacagggtg gagcgaccat gctctggccc
24601 tagtaggttt tttttttttt taagtttttt aaatgaattg gattttttta ggtttataaa
24661 tttctgtctc cccaccctca cttttaggca ttctcatcag gaagatggac aaaaactggt
24721 tagctgtata aaattccaaa tattttcttt atgacccaaa acactctgtt aatttttaag
24781 tgttctatga tttctcgaaa agaaggttca tcttcttttt tttctgtagg gcaagggaaa
24841 cagtgatata actactaaat tagctttact cattgtgtca ttcaaaaatt ctatgcctta
24901 cttagactta actctgtaaa ggctttttaa agttagattg agactctaat gctctgtgtt
24961 tgcacaattt ccccttttgt ccaaggattt gttttggagt gtttttgtgt ctcttgggtg
25021 acacgaaggg ctgcaactgg catcgcgctt gtgggtgctt cgggtacact tggaaatgag
25081 tggcttgggg gtctgaggaa gtggccccc aaatctact tagtctctggt atctctctgt
25141 ccaecggagt cctctgtcat catttttaca tctgtgtgct tataagttgt atcagtttgt
25201 atttttatct tctgactga gatctctttt gaacagagt caactttat ataattatct
25261 gatgtatttt atctatttat ttgttgtatt tctatgcaca ccgacagcaat aggtctgact
25321 tggctgtctc gggtaagaag cctgactcct cgggtgtctg atggccctgg ctgcttacct
25381 gtgactctca caaagtcaaa ttaagcctcg gtgtcttatt gtgtgaaatg ggagagatca
25441 cagctgtgac cctctcatct gccctgtaga ttaagggcat tacaacatca aaagtgtctc
25501 ccaacggcgt ggtagacgca cctgtgagc atcccgggcc ctacgtgggg ggggtctctc
25561 tctctgtctc gcacagatc cctgtgtgct atccaggtc actgtctctt cgggctacac
25621 acccactgac accgcacgga aaggacgtcc cagggttggc tgcgaagaga agtggagagc
25681 gcaagagagc caagctggga cagcagcgga ggccagagga gaggagggca cagatttaat
25741 acggggggag aagtccaaag gagggtcgca ttaaccagca gtttgtgatg ctttgacatc
25801 gaggcttgc tgactctgga gggaactggc cgtcaagtgg taaccaact tgaagatg
25861 gaaaactact atccatgagc acatttttca gatacaaaac aaaccaatca gccccatgt
25921 cccagccac ctctttatc agaattcaea ccttagacca ctatccctc aaactggac
25981 cccagaaaac taactcagga cctgtccact cccagagacc cactggaatt aaactggac
26041 gccaaactta agccagctca cctgtctccc atcggttctt cctgcagaaa caactggac
26101 tcaggctgct gttaaccaaa Ytccctctgc cctgtgacca cccctgtgct cctcttgtt
26161 ggcccctgtg gcttggatg tctcttctc ttgggatctg ggagtacag actctttca
26221 gtgaacataa gtaatggcca catttttaa aagggggta aagggagag ggttctgtt
26281 tctttaagga aactcagggg agccagggag atgtgtctg gttctgcaa acgacacac
26341 tctccaagga ggcccatgcc tcaactccag cccctcact cagaggggtt ggagggagc
26401 cgggtggccc acagtgcact caactttcca tagctcagga tagtatacg cccagagacc
26461 acgttgatgt gggaaccaa ggaactgtca ttgttagatt catgtacct aaagaacaag
26521 ggaggctca cggggaaacca caggccctag gaactgtctt gctctgcca atggatgag
26581 ccccaactgc tctttctct tccatgctt tccactcagg cgtcaccatc aactaggagc
26641 caacttgtga cccctccctc attgcccct gccctcgcc ttccaggt cgcaccaca
26701 cctgtcgggg tcaactgtgc ccttctcca gcaacccca ccaacagccc agggWtccc
26761 caactggggc ctgaacatgg aaactgggca gggaaagcag cccatagctt cgttagctt
26821 ccaactgaaa caaaactct ccttcaacag tgaactgggg tgaatggacc cggcgccaca
26881 ggacgggtgg ggaccggccc tggcagaagt caactcaagc atctagctt cgtgtgggg
26941 attactacc tgactcgcta ctggaaaatg ggggcycttc tgaagccca accagatata
27001 atgttttaac gtgttaataa aaagctcata ttaactgcat cccaattgtc tttctctgt
27061 ctgttttgat ggctggagct cctgcagtgc cgtctctgt atttatca atcatattc
27121 tacaagtgtt tcaattctca catttaaaac caattattct cagtgtgag cttgtgctt
27181 agcagatata aaagggaacc gtggcaaaaa aacagggggg tgaactcccc ttctcagaa
27241 cacaagcagc aagacccctc agaaaactag gtacggttat cccgggtctg agctcagc
27301 gggcttctct cgcactgagc cccacccctc tctcaactcc ttgttgtt acccaaac Yactctctc
27361 ccttagcgca cccccttcc ttgacactcc cgtcactcc cgcagccac accgggctct
27421 cgagcccgag tcaagctca ctccccctg gaaggccccc cgcagccctg agccgcccct
27481 catcaagcag caagtgtgtt attctggag cccctgtgt gccctgtgg cctggcgagc
27541 caccctgcat ctccctgctt gcatgccag agccaggctca agcagcact ggacaggact
27601 cagggaacc cgtgaaagg ctgaagccat tggcaccatc aagaagtggg gaaagcagga
27661 ctggctgtgt gcagaccag cccaggggct gctgacagct cccgacatgK ccaagacata

27721 cttcccgagg cctgctgcc cgtcaggcca gcacacact cctgcactcc aatctgtaga
27781 ggcctcatct gagcaacagt gattgctacc tcaaaacagt ggggaaggcac ctcaggaaac
27841 cYgaggggtc caaaggagaa tgacatcatg atggatgggt gggacaggcat acaggcgctg
27901 caagacaagg atccccaggc ccgctggcat aacaaggatg gacacaggatg
27961 tggggcgagag ttcatgacat gtggaccctg ctgagggtac catccaggccc aaacgcagct
28021 cctcttggtca actgctgccc agcaaatcta catctgat tg taaatgggct ccccacaact
28081 cccacctctc ttctctatag gcaagggaca aactcatcaga cccacgggag cgtgctggt
28141 caccgtggct ccgggctgta gggtctggg gagggtgggt gggacacccc ttgaggcgct
28201 gagggttatc ctctagacc ccagtgaacc atctcgcccc tgtggctgac aagatggagc
28261 tccgagggagg gaccccaggt cccaggaaag cctggcattg tggctctgac cctccgttct
28321 ctctgtaaaa tgaggatlgc agtcccccca tcaaaaagct gtgtggggg tgaacggtc
28381 atcgagggtgc aatggaaatgg tgctgtgca aacaatgaaa cgggttttca ttgctgttaa
28441 accatagttgc cctgcagctc cctgtgaggt cctgggggta ggggtatgac ccttaaacca
28501 cagaagtctt tgcccagctt tgagggtcag actagattgt ttatgccacg ctggcagcag
28561 agccttaaaa gccagttttt cttaaagagag aagggggttct tghtaatgtc cacagattgt
28621 ctttctctat cagcacaagg cttggccagc ttggctcatg gagtctgacc gagggaaggt
28681 cccaactgac ctgaaggagag ccaaatagga gagaggagaa agggccaggga tgaggggcca
28741 ggccttctct caacctcaac caaacctcca caggctatcc aggtctgctc ctgggacgtt
28801 gaaggaagga aatgcaaaa agcttgatcc tggctgctgc ctgggacgtt agaggggaca
28861 ggttctcagg agtaggagtg gcttccaggg ccaactgggt gctcaggctt gctcaggctg
28921 gagagggtgc cccagctccc agctgtgtgt agcagtgccc aggggtcccc gcaccccaga
28981 gacacatggg caagatagtt ggggactggc ttgcaggttt acctgtgctt agggggcacc
29041 accaggcttc agggcaagcc cagggggggca ctggtccata ggagagaggag gcaaggctgc
29101 ttgcaaggct actgcagagg ggaactatcc ctactgcccc caccaggccc tgtgctgcga
29161 atgggggctt ttgggaggtt attaggggta gatgagttca tgggggtggg accctcatg
29221 taggggtagt tgcctatagg aaacagaccc agagaggctg cactctctct cctgtccgtg
29281 gggagaccga cgaagaaggc ggctgtctgc aaaccaggaa gagaatctct actgggaaat
29341 gaccagctgc gcaacatgat tctgaaattc ccagttctca gaaactgtgg gaaatgagc
29401 ctgtgattta aaccctggg tctatggcat ttgttatatc cagcccaggc ggaatgagc
29461 aaaggcttta ttctcgggg actgagccc actgagccc cctccacagg ccagactgca cctgtctt
29521 taactcgggg cgtagtggcc gaga tgagggg cagcacaga tataggaac aaagaaaggc
29581 ctggccagcc cagctggctc actcctgcaa acctcagact ttgggagccc agggcggtgg
29641 atcacctgaa gtcagagttt caagaccagc ctggccaaca cgggtgaaac ctctctctac
29701 taaaataaca aaattagcca ggcagccttg ggcagccttg taatccagc tgcctcaggag
29761 gctgaggcag gagaataact tgaaccagag aggcggaggt tgcagtgagc cgagatcgct
29821 tcatgtgact ccagctggg agacagagca agactctgtc taagaaataa aaaaataaaa
29881 aagaagtga aagaagtgtc atttcccc acggttgagc ctaagcgtgt aatgltgac
29941 ctccaccgtcc acacagagcc tcaagcacag cagggcccc tcaacagctc tcaatgggac
30001 ttctcaaat aaactgacc tggctctctg tgcagcccac caacacctg gcaacgaat
30061 ggcctggggc agcctgacct ttctctctct tcccactggt gctgaggtg gacccagca
30121 gaccagcag aaccaactgc tcaacatgaa ttgtgctct gagttagatc gggacagccc
30181 tctctccagg aggcggagag ggaactcag atgctacag accagctcca gggacacgg
30241 cctcgaggga aaggcaggga accagagat atattagttg gaactgccc taatggggct
30301 ggaagagcgt ggcattgtcat gagactgag aaacagaaaca gggtagtca gcaagcgag
30361 gctggcgctc ctggcagagg agcagcccca tgcaaaactc ccagggttgg aacagcgag
30421 gcaccttcca agcacagagg ggaatgtgaa gggatcccc agcacagagg ttggggagg
30481 agcactgagg ccagaccacg aagtcaggag aaggattcca aatctctct aagacagtg
30541 ggaagacatg gaggaaagat gatggcgagg catttaagaa gcttaactca ggtgcttac
30601 gaaagtgggg tgaagggggt caagatgtgg gaattgtggg ttaagctttg ggtgagtgt
30661 cagagaaggg acggatgtga gggatgatgt ggaattgaaat aggggtctac tggcccgct
30721 gagcaggagg aggtcggggc tctggaagg cctccggca gctcccacag ggcgaggga
30781 ggaagtctgg ggtcggccct ctcttggtc ctggccaggc cctgccccct aggaacgaag
30841 atcttccacc ttctctagcc ctaccctgg ggcogaaggt ggcagctctc atctcgaag
30901 ggaattctgt gtcttctctg acattagaca gatctaactc tgcctcctga tttggagacg
30961 gaggtgacca accactccaa agccccgtc gctgggcatg gaactcctg gtcagacccg
31021 tcagctctac attctgctt gagggtgggt cgaggcagct gagccctgg gctcagct
31081 gctagcgtgt gcccctggc ttgggagctc caactggggc ctgtccggag gtagcagctc
31141 ccaactctctg cgtcgggtgc cctgaccctg gctctggaa gctcagat gttcggaat
31201 taactggggg tccctggctg ctgaggaggt aSagagggca tgcctcga tttccagc
31261 aggtagccag ggagaaggag gccaggcgcc agccaaggcc cctcagccc tccctctct
31321 gtacgcttca actgtggtgc gctctccctg ctctctccc caggaaacct gctgctctc
31381 ctctcttcca caccagcaca gcacatagtt cccaggagat gcccaactc tgggagagc
31441 gctcagacaa ggccgtgggt gggcttgca ttggctctg ggtcctctg ttggtgctg
31501 ggcYtgggat cctgctccgg agccctcaac ttggccccc cccaggccag gctctcagcag

31561	ccccctccag	cagtcaggaa	ccccatgggg	cactgtggcc	cagagaccac	tgaccactgg
31621	cctgggtggg	tctcctgccc	ctcgtgtgcc	catggccaaag	tcacacccac	agtccggggg
31681	cggggtcctt	ctcgtgtcca	gcctctcagt	ccccgtagggt	taatatgatga	catcagggaac
31741	agacatcggt	ggcaggcgagg	gtttgtgggg	catgtgagag	gcagggggtgc	tcacagatgtt
31801	ccttgtctctc	cgtagggcctg	aggagtgagc	cccgctcctgc	cccccttcacc	atccccagggg
31861	tctacctgtct	tttcaggagg	ctgcaaaaatt	gccaaaaattt	ctctgcccaagc	cctggccaga
31921	ggcctgcgag	gccccacttgg	gcatgggaac	cactgggaggt	aatcagtgta	attccccagaa
31981	aaacacatggc	gtcctcagta	aggacacagc	ggcaacatatt	ctgcagagagc	gagagtcgcag
32041	tgatgtggca	aaaagagcaa	ctgaagacag	aaggcgacat	ccacatgcctc	gcctcttgca
32101	catccccctc	ctcccccggc	ccaccccact	ttctctgcca	gacagggagac	aggcggtagg
32161	gctccccaac	ggctccccctc	gtgtcctccc	caccccccca	cccacacaca	ccattcttca
32221	ctacgatcac	caagcgggaat	agatctcacc	ttccccctcc	actgtctaaa	accctcaacg
32281	gctctccagt	gccttcggaa	gcaggggccca	gccctcgccc	tggaatctgg	aggccctctcc
32341	acgctggccc	tggtgtggtt	ctccatgcgt	caettcagat	acctccccac	atggcacccct
32401	tgctcctgac	agcccagcca	gccttcccaca	gagccggcgtg	ccctctcctg	ctgcctgtgc
32461	tttgggtccag	gcctggaaatc	ctcactctcc	ctgtgcagcc	cctcaattctc	ttagatatctc
32521	caggagcatt	tgaaaaaaaa	gactcatcct	aaaaataata	ctcaaaagga	aagagggcagt
32581	gaSctccttt	cttaaacctc	tctaccccca	aggcatgaga	ctcgaaagtcc	atggcctgacc
32641	tccaggccag	gaggggtctgg	ccctgctgag	ctctctcggg	ctgcctctgc	ccctgctcccc
32701	gtgctccagc	gtgctagcccg	gtgctcctct	gcacaggcgag	ctgtcttcca	cctcaggaaac
32761	tttggacttt	ccgtgcaactc	gtcctccagca	gtgYctagcc	catagtagtg	gctcagtaaaa
32821	catttgtctga	atgaatgggR	tgaaatcaact	tgaaagatatg	gaataaacagg	aaatggcgcaa
32881	aatgtatagc	ccagcgtagg	cctgtgaccc	ctctcatggg	gagggccccca	acaaagggcc
32941	tccccctccc	tccccccagc	agccctgggg	gcctcctccg	catccaccctc	accagggggg
33001	tgctgtgtct	gggggagaagc	agctctcctt	tcccaggttg	cagRggggag	gggaagccagc
33061	ggacaatgct	cttgagtact	gtctgagaaa	ttgatcctgg	aatcttttaa	ccctgggggg
33121	aaatgtctaa	agccgcagcc	atctctgtgm	ggagagaggc	caagggctctg	cttgtttggg
33181	gaacacacatt	gtaattgcgc	ttttgcttat	cttctgctat	agccaaYgc	ctgtctctgga
33241	atagacaatt	aggaaggagc	ggctctgcatt	ggcaatgtaa	agtttgccctc	actggttgga
33301	gagttctccac	caggggatgg	tgagccctgt	ggcgtgctgt	gcggctgctg	ggcagaggtg
33361	actttgtcct	ctgcacaaagt	gggggagggg	tgacagaaagg	ggggaaacggg	tgacaggtctg
33421	cgtgggggga	ggcctactgg	gctgcacggc	acagagggaga	caggcagggg	ccctgggaaca
33481	gcagccacccc	caaccttcag	gaaagtcccc	atRctctcg	ctcagctgag	cctttggggca
33541	ccagagggcag	actgtggccc	accctggcaaa	accatgtcca	gggacacggc	aggcctgaaa
33601	agttctgtcag	aggcctgggg	gagggggggcc	tgctgtctctg	tggtctctct	gggacccccct
33661	gaataaagtg	aggagaaggc	tgctctctct	tgctctctct	ggaggccagg	cttgaggacac
33721	caggatggag	gcagagaaga	gagtgacggc	aggagcaggg	gacagaggctg	tgctgggcctc
33781	cgtgggggaa	gagtttaggc	attgcatggt	gcctaaacgg	ggctggccaga	gctgggctct
33841	atcatgagaa	tcaacgttcc	cattgcagat	ctctctccac	ctcggtctctc	cagcagatgc
33901	ttcagaggga	atgctagaag	aaacgtgaat	ctgtctcaaa	atatacaaat	accatgaaNc
33961	ggaaattggcc	aggatcagaa	ttactaggct	cagttctatac	atcaaaaaagt	gaggttggtca
34021	gagtcctctc	cttcaatacc	cagcgcaggga	acactctgcc	caaggccacc	cctctcaectc
34081	tcaaggcgag	tctcctgtgc	ctgtccacac	ctctgtgata	cctctgcaga	tgctcaaggg
34141	tggggacggg	gctctccac	ttccaggcca	gctcatcaat	gcacagagcc	cagcagggtgc
34201	tcaagagatg	tgctgcaccc	ctcagttgtg	cactccagca	ggcagcctga	gaagacaggt
34261	cgcattacaa	accctccccg	gtt tgaacta	acgacgggcaa	gaggtctgcc	ctccccctct
34321	cctgaatttg	gggggttagc	agagtgtagg	ggggatatct	ccactctgag	ctgacatattg
34381	atgggtgcga	gcattgggctc	tgcccaacgg	ctgctcgctg	actttggggcc	agttacttaa
34441	cctgtctggg	cctcagtttc	tttatctata	aaattggagat	ataaaYagt	ccctgtgtctc
34501	agggtgttt	caggttgaag	gcaaatgaa	aatatcacg	tgctctctga	tgaagagatc
34561	accaaaagg	ctttgtgtga	gcaataaagg	tttttaatac	cctgggtgca	gggtgggctga
34621	gtccaaaaag	agagtacga	aaagggtgggt	ggactaccat	ctgtgtggat	ggagctggga
34681	tagaggtgg	agttaggagc	aat ttttttt	tttttttag	acggagagct	ttcctgtgtgc
34741	ccaggttgga	gtgcagtgcc	atgatctcag	ctcagctgca	gctccacctc	agggtgttat
34801	ggcaattctt	tgcttcagcc	tcccagtagt	ctgggggtcac	aggcgcccgc	cagtgccgctc
34861	gggttttttt	tttttttttt	tttttttttt	agtagagagca	gggttcaaca	atgttagacca
34921	ggatgggtctt	gatctcctga	cttcacatgc	tgccccctcc	ggcctcccaa	agtgctggga
34981	ttacagggct	gagccaccgc	ggccggcgag	gagcaatttt	ttgtggggat	ggtagcggat
35041	ttaccaaatg	catcttcaag	ggcggaagaa	tattacaaaa	tatcttctta	aggttaggggg
35101	ggacaatatt	acaaagatct	ttcttaaggga	tggggggtggg	gaagaatatt	actaagtatc
35161	ttcttaagtt	gggggggagag	aatattacaaa	agtatctctc	caagggtggg	ggagggtgat
35221	catacaaggt	agattccaaa	ggcgggggcg	ggcgggggcg	ttgggcaata	ttaccaaatga
35281	cattacccca	agggcgggga	gggtgttatg	ttcaaaaatt	caattgtatg	atacagtatg
35341	gtggggcgag	aacagatcac	aatgggtggaa	tgccatcagt	taaggcagga	actactattt

35401 ttcactttctt ttgtggatct tcagttgttt caggccatct ggaagtatat gtgcagggtca
 35461 caggggagat gatggcttag ctggggtcga gaggccgtac atrectgret tctatatata
 35521 ataagaaaaa caaatgaa tagtgggtga gtgtggggg ggtgaaaaat ttgggggggtg
 35581 atatggagag aatatgggtg atgcttctca gggctgcttc aggcaggatt aggggcagca
 35641 tggaaaaccta gagtggggaga gattaaagctt aaggaagatt ttgggggtaac aggtgatatt
 35701 atggggtggt tagaaggagc atttctcgtg tagaatggct ggtgatggcc agaaaaagctt
 35761 ttggtatgaa ttgaaaaact aaacgggaag cacagctgcc gaataaaaag agaaaaagctt
 35821 gtattaaaag actaagaatt gggaggagcc agaacatcca attagagat gcccaagggg
 35881 gtacgacata atattttgtr ggtttggcga gtttttgggc tctatccttg agttatgtt
 35941 gtcataacac agggccagact garttagtga aaaaacacac ctagggtatt cggccaagctt
 36001 attgtctaca caaagctcgg ttgtggtctc ttacacagga catgctgtag aaagttagaa
 36061 tccccagcac ctagtacagg ctgaatgaaa ctgtcatcaa attacaggag ggtccacacg
 36121 cccaagctgt ttccctggga ccccatgcac agaattggtcc acctgggctt ogccctaca
 36181 cccacctgt cctcctcct ggcctcacc atcgggagag gtctgagtag acagcttgca
 36241 catgcccctg ttgtttccag agctggaggg gctgctgtgc ccaggacatc ccgcagatgc
 36301 gaccacgccc tggagctctc ccgctggctt gcgtcgctt tggggaaaca acagccttgc
 36361 caccacttag ggaactgaact tcccaccgcg tgaccggatg cctcatcaga aagggaaatg
 36421 ctgcagaaag gggaaaggacc agtgcgtcgt aggcgcagc agcagaaggc acaggccga
 36481 gggctcaggc cagagaaccag ttctcccacc gggcctctct gccactgcga gccactgaca
 36541 gggaaaaggcc ggaataagtc agggattcttr atcttctga tgcagtggctg ccaggacggcc
 36601 ctgctcctct ggaactctgc tgaagctgtc ggtctcctc ggcactgcga cctcggggg
 36661 ctgagtgaga agtttctctt caaagaaggg ccctaaagtgg ccagggtctg ccccaagccc
 36721 ctgctcggca tcaactggag ccgctgtctc agaggtctgt tggcccgccc catcccgcta
 36781 ctgagctcgg tcaactggag ccgctgtgaga aacttgagac caatcgata
 36841 ttatctcccc acacataata ttctcgctc gctctgtcac catgtgacc taagaaaaag
 36901 gggaaagaaa atgcccctat ggaaggagag agcaggagga ggtcgtgagc cccacaccag
 36961 gggggggggg ggcgaagca gctgcgggg tgggggggg cccctcgctt agagtgagat
 37021 ctggagcgg ctagcctatc cctaaccgtg ccccgctctt aggttgagat ttagtagtgc
 37081 cccggggttc tggctccac ttcccactga gatgagacc ttctctgagg gtggagggg
 37141 acccaactgg ttggcgaagc ggtggttaga cagaagaagg atctcatggt aatgaaatgg
 37201 ctccaggcag cacatggatg cccctcgctc ttggtgggg tgtgtgggt gtagagctc
 37261 ctgctcgggc gggccagggt gtgggtagct ggcacttgct ggcagagggt gtagagctc
 37321 caccatcaa ggcatagggc ttggatggg caggggtcca gcatgcctaa gtggcccgag
 37381 cgtcagagcc gaccactca gagcaaggct gagaggagct caggcagcc tctcatctc
 37441 ccaggctatc atgcatgatg aggaagaagg tagagcccat cacttgacc ttgtccctc
 37501 ctacagagat gagaaggcat gacagaaggc agcgagggc aagagtgaag taaaggtctg
 37561 gacctggcca ggggcccaca aggcactcag agcgctcgc tgcctccacc gttccctc
 37621 aagaagggtc ctgcattaca agctcaagcc aacaccagga cagatggtc cagcgctc
 37681 ctgaagaagc agtcatgcag ggtgctttcc ttgtcagcag gtccaatcc ccttgacaa
 37741 tgggcttggc gggaaacggc ttgtcccttc ttgcagact tggcagatg aagatgaga
 37801 cagagatgtc ttatgtggc tagctcctc cagcagact tggcagatg ccgaattatt
 37861 gaaatcctgc ccaccgctac acaggaatac cagaacaaa ggaagaaggg ccgaattatt
 37921 ggaagacaaa ggttcgtaga tcaactctgc ctggagcagc aaggcgtagc atggcgagc
 37981 cccagggcag gctcctcgc aggcgggggt gggaggtctt cagagctgca gccaggttaa
 38041 cttcctctc ttgtgtgca gggaaagttc cttgtctg tagtttctc actctgta
 38101 catctcccag gggcacccta tctgctcctg ggtgctgtc catcctgga acttctggc
 38161 accatcgtga ttccagatgc cttcgtttg tttagctgt tgcgtctg actagcagg
 38221 gaactgtctc cattcttgt gatgcaacct ggtgtcag tactcctcgt tttcccag
 38281 ctccaggacc aaggccaagc tcaagccagc gcgctctc acccccag cctctccag
 38341 ggtggccgag ggtggcacc tcttttgac aagggtctc tctgttga atgagccct
 38401 cccactttcc acttgggaga caccactc gtgaccaaaa cagctcaag ctgaccoca
 38461 ctgtgactat ctccaKacc aggtggagac agctgacgt ctctcttggc ttgcaagRcc
 38521 ctgctctcRc tgtgcgaagg caagcaggtg gcccccaga gaaaaaagct tctcaggag
 38581 gggagactggc tggggctctt cccagaaatc atgggcccag tgaacaaact ctgcctcgc
 38641 tggctcggctc tgcattctc ccctaYgcc gagtactcc cgtgctgca tatcttccc
 38701 tccccacac ctgcgagtga gtgttatccc cgtgtgcca trtaggaagg cgagggcag
 38761 agagatttag aggtcacat aagtgggcat gggagtcagt acagagctc tggtagcaaa
 38821 ttgtggcaac ctgctgcaac tagaaactgt atccccggac ctgcccctc agagctcagc
 38881 atggcagctg gagacacagc ctctgtcara gaagcggaag cataaagaca cctccacagc
 38941 aggaagtcca caccacaga tagagtctat tttratggga gtctctgaga tcaacagctc
 39001 ttatgtggaa ttcaatttagt cactattcca ttaaaaagag attacttag cactactac
 39061 gtgccagccc ttttcaggat actcagtgaa aaatcaggtg ggcagaccct cctgggctc
 39121 garcaaaagt cacccttggg ggagtccta accctgagga cacacacac caagagacc
 39181 cctggagctg ggaggaagac ccaggaattgc ttgactgcaa ctacagctca ggggcagcct

39241 gggaggggcc tcaggggagt agagccacag ccagggtcag ccgacagccc ggcgtggcac
39301 atcatcgccct ctgttcttgt ggttcagctg tcaacctcct gcttgaggct gcacaagaca
39361 agctccagag ctgctcaag ctgtcatccg ggtgtgtgga tgggtggccc tctcgtcca
39421 cagggaactct aactccatt ctgacaacca ctgattccccc gggggaggcc aaggacagga
39481 tgcctgggagg ctccacatc agctatcagg aggtcgcggg agggaggagg agggaggagg
39541 tggggagagt tccactggg gggagggttc tgacatttc ccacatttc ggtctgtcgg
39601 tccaccacca agacccctgt tccagtgctc cccacctgct caccattct ggtctgtcgg
39661 cctccccagc cctgctggaa catgtgcttc ctgcggtctg ctgcccgtgc tccaccagac
39721 atccccagag cctagggtcg ctctatccgg gtgcactaag ggagtgagtg ggcaggcggg
39781 Ygggctctcag aacaggagga cacagttagt ctatgcagaa gccaaaggac agcaaggagg
39841 agagcggatg gctctggggg agaggcgggt gtggccaggg gactccaaat ggtctgggga
39901 ggtcagtgag gctcgagatg agaaccYgcg gagcatgcta ggggattctgg gagctctega
39961 tcttcaggcc agacagtagg gaggagaccag ctcagctca gtgacctgat gcaggcaagg
40021 ggtcagtgag gctcgagatg ccaatgtctc tgaggtcaca ggtgtgctct ctctgacagg
40081 gaacctcaca gggccactgg ggcacaggtag tccccctct caggltggga ggggtgttct
40141 ggcagcagtg gcaaaaccag caaatgtgag ctaagagaga ggagltctat tttgtagacc
40201 caaggtataga aagcataaag cccaaaccct ccaaaagct cccaaagct ccagtctcgg
40261 acaaatcaca ctgcctcccg ctctccaaag cgcgcgcgtg tggccaccac aggagtcctc
40321 cactgcgatg actgagtgag cctgtgcttc cctgtgcttc aggagctct cctgaccagc gcccaccagg
40381 attagaggca actgcgggtg ggtgtatcca cgtgtcctga tcccgcctccc ctactctctg
40441 accttctct gatgaaccaa ggtgtatcca gaggcaggga gaggcaggga ggcgaacct
40501 tgcctcfagag tccccagtg cagggatggg tccccccaca cccgtgtctc actgagcact
40561 gactgaggag agctcccaat cccagcatgc tccccccaca cccgtgtctc actgagcact
40621 tctctgttgt tcaactcgag tgggtgctccc aacagcccgg tctctgtctg actgagcact
40681 caggggcgag caggacaggc agtcgctccc tctctgtctg gctgggaaga acactcaggc
40741 caccacaca cagacactgt cctgcagcgg tctctgtctg cactgttga gttgtgctc
40801 ttaatttcaa tctgtggccc acattgaaga ataatagat tttccaaaaa aagtagattt
40861 cccagctctc attgaagagc gactggccag cagtgtctg ctcttctcta ccaataacata
40921 agcagtgga accacagga gctctgcagg ggggcgtggc ctctctcggt ccccagctc
40981 ccaacctct caacagcctt cagccactcg gggccctcgg gctggtctac tgtgagctg
41041 tggggagct ggcctctctt gagccttgcg tgcagctcg caccacccc tggccctcat
41101 ggtctgtctc cctgtgtgga cagcgggcca ggtctgaagg ccagcgtagg aagcgtgagg
41161 aacagagcaa gtgcagcaa gcccccagg cccacagcac cactagcaga gatgtcttc
41221 cgtctgcttc ccccactta agggacctg aggttgcga cagttcaga cgtgtcttc
41281 tcttccag gaaccagagc cagcgactcg tgaacttga tctggggcga cgtgtcttc
41341 ctggtctgtt tgcattctg cggcagattt caccagact cctgagaaat cgtgaaag
41401 actctgtcc gaccaactgc tccccaccg attgtttga acccttggc aggtctgtg
41461 tgggtgaaga ttgagcccat atgctctcg aagtgaaga aggaataata tctgaatct
41521 aagtccaggc caatttctgg attcaattag aacgtctga gaacatact agacagagt
41581 ggcatagaga gaattcatt attcaaaaa agcgtctga agcagctag aagactctg
41641 gttctgtgc tgggagggt caaagggtat agccgcgact agcggcgag agagcgag
41701 gattccagct tctccatagg cctaggagct tgtctgccc cagcagaaga gatgtctga
41761 gggctcagg attctccag acataagca gttaacaga ctgggagaga ctattctgga
41821 tggccggaaa cctgctctc ctgctctgcg gttctacag ggttaaga agttcattc
41881 atctttaaa aagaatcccc gaagtgcgt tgcgctcgt ttgagctgat aaggctagt
41941 tgggggtggg gttgggtctg agcagccac tgtgctcgt attccctaa cagtacacc
42001 cagctgcact gatgagtga aaggcacatg cttgctttgg aggctctggg ccacacagc
42061 tcaaggcagg cactaagcaa gtacatttct gcacagcctg gttagcaaaa tctctttagc
42121 acctatagg gaatgtctc agctttgtgg gccatattg tctgttga actaccaca
42181 ccaactgtgt agcacaamaa tggccacagg caattcataa atcaaatagg gtcgctgtg
42241 tccaataaaa ctltattac aaaaacaggt gaagggttat agcggcgag gtggcgagt
42301 tgccaacaca gctttagaat ctccagttta ccttgacaaa ttcacgtgaa tttgtcttc
42361 atgcaagtaa tgcgtccaaa ttacttttaa tggtaaaa tttgttttaa acttacc
42421 tccaaaaagg gggccacttc agggggagat gYttacatc tgtctgtgtg cctctggcag
42481 gctcgagagt acgtgttctc ggttagcaata ggttagcgtg gttgctgtg ccagctggg
42541 tatcaccaca tgcaatggtc tgctctctg ctacaaggag tgcgctggg ccaggggctc
42601 acaagattca tcaagggtc ttgtagatcc tgatttttaa gcaacacctt gaggcttag
42661 gctctctcgg aagagttaga ttgaatttag cgcttttagt aaatgctga agccaaagt
42721 gaagctctag ccttggggcc agcttttaa ttcttgaag ctctggaca tgcacatgt
42781 acttttaaaa atgtatctgt agccaggcgt ggtggtctg cctgtgact ccagctact
42841 aggaagctga gcagcagga tccctaagc ccagggtgat gactccaca tgcacatag
42901 catgaacctg ctctcaaaa aaatcttaa aaaatgagt aggcatagt ggtgtgctc
42961 gtgctccag ctgcttggga gctgaagtt tgaggctga gttgactat attgacgat
43021 tacgctccag cctgggtgac tgagtgcagc cctgctctaa aaaaagaaaa aatgcata

43081 taacacatatt acttccacaa cggctctgtat cttccaaagcc attcatccct gcttcccat
 43141 taacttgagcc tgcagaagag tccctgtctg taactgggga cgaaccttga ggtgccacac
 43201 acatcttgagc tgctctctag catggcagcc tcgctgttca ggcaccaagc cactctttag
 43261 ggggtcccgag gccacacac atggccacag catctggttt gtgagcagag actagcccgag
 43321 gcgctggcgag agcaagtttt tgaanaatgtg agaaatacca aagccccagc gtcagttatca
 43381 gggctgttcca gcaatgtcac ctgtgtccca gctctggcagc gctctggcagc gctctggcagc
 43441 tgtttcccaa gcaggaaagg tcaaaactatga cagtgtgtct ccaatgccct cggggggcggg
 43501 gggcgggcgtt gggcagagga gagcccttca tctcccgcag tggcccgagg gaggcagaga
 43561 acgcaaggtg agggctgcag gctgagccagc agtgaagaag tggggaggagc aggagcagaa
 43621 tgaaggcaga gggaaataga gtcaactctgc cagtgaaggc gctgaggggc cagcaaaagt
 43681 ggaatgggcca cacatctgcc cacggctctc actccatcct cctagcttca cagtgtctgg
 43741 aaggggagacc cctctgccct tagacggggg aatgtgggca tccccaaggg gctcagcat
 43801 cctctgaccag gctaaggggc gctgagccctc agggagaccag actcatccca ggggttgccag
 43861 cagccaggagg ggcctacagt gcgtgtgtctc aaactccagt gtaggccggg agcagtggtc
 43921 catgctctgca atccccagcact tttggggagg ccccgctctc ggaatcactg aggtcaagag
 43981 tttcagagcca gctctggccaa catggtgaaa aggtacttgg gaggctgagg caggagaatt
 44041 ccaggctctgg cggcacgtgc cagtgtctcc aggtacttgg agccagatc atgcatctgc
 44101 gcttgaacct gtagaggtggc ggttgcaagt atctcaagga aaaaaaataa aaacacacac
 44161 gggcgacaga gcaagactct cgcgccctgc cctgcacag cctcagaca caccagctgc
 44221 gagcctctgt ggacaagggc cgcgccctgc cctgcacag cctcagaca caccagctgc
 44281 cctctctagca cggctggctt cgcctggggc cctgcacag cctcagaca caccagctgc
 44341 gctgggggaaa ggcacacagag aacgggaagt ctaggtggca tccagggggc acatgcaggg
 44401 cagcaggtgt tccctcaggc tatggggctc tagatgacca gggcctatcc cctcagactt
 44461 ctagggaagc gctgaactca tagatgacca gggcctatcc cctcagactt gcatgtgtgt
 44521 gggcttgaacc atgtcagctc tgtcttgaag ttcttgttca tctcactccc gcatgtgtgt
 44581 ttttgaagtg aaggctcctg gggcaatgca gcatgtgtcc atctcccag caggagctt
 44641 cagctgtcccc gtcagaggccc aggatcaggg atgtattccc caggcctatc caggcgact
 44701 tgggtgggtg ccatccggcc caggtctggc caggtctggc caggtctggc caggtctggc
 44761 ccagagcccg tggtttcggca gagggtgaRcc ccccgtagt gacacttcca gacacttcca
 44821 atttgccctg gccataogct ctgggtcagc caggtctgagc gcatggaagg gctgaggaag
 44881 gggctcacttt cccacagacc cctggccagg gcatggaagg gcatggaagg gctgaggaag
 44941 aggaagactct cagacccagc tctcctgtgt tgggtgtgct gctcctgtgt gctcctgtgt
 45001 gggccccctga cctgaggagc ctccccctgg cctcctgtgt tgggtgtgct gctcctgtgt
 45061 ttccctccag cctccctgag tctggctgtg tctcctgtgt tgggtgtgct gctcctgtgt
 45121 acccctcagg gatcagccac aaaaatacaca ttgttttaat ttcatgtgac ccatcagtaa
 45181 gctaaatgct ctaatatatt gcatattaaat agccatcgca atgattaaag atgaataaga
 45241 aattcatgct aataattcaa ggttccaatt ttgtttcaat ttgagcaata ttaaaaaaga
 45301 aataaaaaaga acagcaggtt gagagacaga gacacagaga ctaaggaaga aaggaaagtgg
 45361 cttatagctc cctttgtatg gcaccttctc ctgctcttgg aacaagcgat tttgccccag
 45421 catgtgtgct tggggcctgg gagttaactac tggctctgct tttagggcag tgttcccac
 45481 tggtaataat ctgacaggtt ttgcagcagc catagattga ggggaataaa gaaatgcatg
 45541 catgtataac atttggtcct ttgaaattctc tatctcccgc cacacacttt tatctctttt
 45601 tttttttttt ttttttttga gagagaactc caactgttgc cccagagctgg agtgagtggt
 45661 cgtcacctca cgtcacgca cctccgctc cccagcttca acccatctc cctcctcagc
 45721 ctcccgagta gctgggaata caggcagctg cagctcactc cggcttattt tctgattttg
 45781 agtagagatg aggttttgtc attgtggcca aggtgtgct gaactttga cctcaagtga
 45841 tccgcccgc tggcctccc aaggtcctag gatcacaggt gtgagccact gactcaagc
 45901 tatcttctc ctactcttgg gagaatttca ggtggaggc tgetatgct aaacccaaga
 45961 agctcagcg gttggtgag aactggagct tgggaagac aactggcag tagggcagaa
 46021 actcagaaga ttaacttctg gggaggcagc agctgctma acacatcacc gaggcagaa
 46081 gcttttccc tgggaaggga gatcttccc gatcttccc tctgttcaac cacaacaaaa
 46141 tgtaggacat caaaactat cccagcgga aatgaacaaa agctcaaca caggactggg
 46201 gcaataattt cccacacatg tgacagacaa aagttttat tacttaaga caacagctc
 46261 tcaagtgaat aaaaacccc caaatcccag cagaatcaag ggcattgaat ggcacagaa
 46321 aaaaataaaa taactctgtg ttaacaatac acgtatatatt tcaaatcag cctcaataac
 46381 atcaagaaaa taacaactta caacaaaaaa acttatctcc tgtgttaata aattaccact
 46441 gctcagaaga actcaaggtt gtaggagagt ggggagacag tcaattgctc aatgagcgg
 46501 gaatagatga aaccataaca gagaacttcc cagtgtaga agcttaaaaa atgggcatc
 46561 cgtgtggcca gagacttcgc tttcagcaac tcagcttaag gaaataagta tgcacagag
 46621 taagatttcca ctctatgaac gtgatacagt tagtgccagt tctgacctc aatggagaa
 46681 taagttttag tatgtgaata tgtcatataa ataataagt aggttaaaaa tgtgtgacac
 46741 tcaratgggt acatgttatg caatcattaa aacaacgtag tagacaatta aactgtggga
 46801 gaaagcattc caactctac aggaatttta gctaatttta catagYaana aaatgctccc
 46861 ggcacgggtg ctcacgctgt taatcccagc actttgggaag gccaaagggtg ggtgtcactc

46921 gaggtcaggga gttccaagacc agcctgggcca acatgggtgaa accactgtct tttactaaaaa
46981 tacaaaaaaa aaaaaaaaaa aaaaatagac ggcgtgggtg gtgcatgcct gtaatccag
47041 ctactcaggga ggctgaggga ggagaatcgt ttgaactcgg gaggaagagg ttgcagttag
47101 ccaagatcac accactgtac tccagcctgg gcacacagat gagactctgt ctcaaataa
47161 ataaataaat aaaaataatc atctgtgtta aatatggact atttcaataa accaaaaaa
47221 argcttttct agatgtaagg aggaagataa gaaaactgga cccagcaggg gctgtcgttag
47281 tggggcagggt gggtggttgc taaaacaaag agcccttttc tgtgccaggg tggcctcatg
47341 aagtgctcca ttccaagacc gcttcccttc tggcaagtct gagatgaaaa gggcagcttc
47401 atctttccca tagaaaaagg aatatcaagt cttgtcttta acatagacag gaagcagccc
47461 ctccagrgat ttccactgct cccggagcca caggcgggtc cctccactgc gcgagactg
47521 gcggcgggcg gcttgcggg ccccgacttc ccttactgtg gccctctgtc tgacaggcca
47581 ggctgecttc attttccag cgggcccact tcttgctca gggcctggc gcgctgacc
47641 ccttccccca aacctgggac aggcgatccc tatcctacct ctteccacat ccttccgttg
47701 tctgtagcct cWgttccctg gatatactca gtgcaccagc ccttgagctg tagaaaagctt
47761 cctgctctcc cctacccttc ttcttctccc atctctaac tattttatt ctttcaatgg
47821 atttgcaat atttgaagt cctcgtttg ttttcaact cagtactgtc gcgctgccc
47881 ttccagaaat gcaaggttag agggttacgt cctgtcatgc tcaacgtgac acccttgccc
47941 tgacataata tagacgcccc ataaacaccg ccccctatct cagcagctga acacataaca
48001 agacagaggga ggatggctat gcccgagaa caccctatct cagcagctga acacataaca
48061 ccgcacaccg tgtttctcta ctcaacacat ggcgcggggt gggctcgagg ggcccgaggc
48121 tagtttgag ctgtacctca tgggacacat ggcttccagg gtttccgtgg caggggagac
48181 agagatggag gggactccaa gctcttacca gcttgggtcc acagtgacc caggctcaet
48241 cctttcacac agccactag gcaaaaactgg ccatgcagct ccacaccagc tacaaggag
48301 tggggggagg gggacacagg ggtgatcgat gaggtgac tgcctgtgac accctggag
48361 cccatctccc agcccaggct ctgcacaaaa agatcccaa agggaggct gcaggcgaa
48421 tgaccccaaa atgatgagg aaactgagg agtgagggg cctcaggctt gactgtgccc
48481 tgcactcaag ctgcacacac ctcaactctt cctcagaa cctcagaa cctcagaa
48541 ctcaggccgc cccacacagt cctctcccca gggctcagc gctgttcca gggcaccagg
48601 tggcgacaca ttcttgcat ccagtcaagt gggcagctca tggccaggg ctggcccaca
48661 gaaccggggc tctggatgtg cgttgcacca gcccacatg gctcagaa cctcagaa
48721 acctcaagg gagccagatc ccaagcgctt cagctgggac tacaaggag agtttgag
48781 ttctgctcca gccaaggga gccctgtgac cagctgggac tacaaggag agtttgag
48841 gccactggcc aggccttaga gggcactgtg ggggttgggt tegaggccc catagacccc
48901 aagtctcagg acaggggcag gtaggcagat ctccatcca ggtgaatga atcaaaaac
48961 tcagcacagc agctgcagc cactctggg accgtaaaaa agaaaaaaaa gacagaagtt
49021 agatctctct cggggtggct acagtgttg tacactgttg ggggggggca aatggccaca
49081 tggagagact agaactcctg tacactgttg ggggggggca aatggccaca gcaactgtgg
49141 gaaacagggt gaaactcct caaaaaatta aagaaagaaat taccctgtgt ccgcataatt
49201 tacttatggc catatagtc aagaaaacgga aaacagagtc tggaaacacc agcttccagg
49261 cagcactatt cacagtgcg catgaggggg ccacccagggt gccatgatgt ggttgatata
49321 acaaaatgtg gccgatccac ccacacacgg gaatactgtt cccctgaaa aggaaggaca
49381 ttcttaacaca tctgtcaatg tggatggccc cgaggatgtc atgtcatagt taataagcca
49441 Sacaacaaa gacaattatt gttgtacca ctatgcag gtaccaagag tagtcaaaat
49501 ctcaaaaaca gaaagtggag tgggtgggtg caggagtctg ggggggggga agaaaggct
49561 gaaactcaat gggcacagtt tcaattttac aagatgaaga Wtctgtgga cggggggggc
49621 gatgggtggc cagccacatg aatgtgtcta cggccacgaa tgcagcagctt aaaaattggt
49681 aagaorgtga cttatgttct gatatattca atktatttaa attaaaattt attaaaattt
49741 tcttgggaca aataaagaat taaattaaaa attcagttct tcaaggacac cagccacact
49801 tcaagtgttc cagagcccg gtggtactg ggcgagagc gccagacga cctctcccc
49861 agcacccggc gaagctctgt cgcacacagg ggtcttgaaa ggttaccctt tctcagaagg
49921 tccccarctg tcaatgcccc tcttcccaa gaccacaaa gggctgcaga tctcttagt
49981 ctgtctctag gactctccg gcattgcagag gctctcagtg aggcagcagg gcatctcatg
50041 tcaggggcaaY gccatgctcc aggcacggcc gcgcttctc tccctcgccc cctctcgctt
50101 ctgtcagttg gagctggggc tcatgtcaga ggcagagacc cccctgtgac cctctcgccc
50161 cggcttccca Yggccgggga ctctctgcca cgcccccgcc cctctgagg aacccccca
50221 tgtgtcccatg acaagcacac tgttaaacag tttgtgtaat gaatcagtgc tcaacttctg
50281 gcaacagggg acaacagaca cggcccaaat acttctgaca ctcttgagag tttgtctct
50341 aaacggtttc tctgtttgta taaaactcat ctcaatttaa gaatttttaa atttctcatg
50401 ctttctcaga aacgaacctt cccattattc gtaacgttgg gtcaggttgg atctcagg
50461 tttgaataaa gttggaYggc aaaagtgaga gaacgttcta gacatctccc aaccagctgt
50521 tgtgacaggag agaccctggg caccaggaga aagttgtgat tgcctctgccc cactcatga
50581 aggactcacc acggggccca cattttgtca gcgcctctg tgggctatca gaagtgaacc
50641 ctctcggact gtcactaaac agttgccactg tcatgataac aaaaacactt tttatgtgga
50701 aactgaaaac gtgttaccac tactgagaga ctgaaagaca cattactga cattatctca

50761 tgtaatcacat tcaacccccca cgtgaaaaac actggcccatt ttacagatg agaaaacYaa
50821 ggctcagagg acRggagcag cgttcccaac gacgcgcctg gaaaagcgga ggcctggagg
50881 YRccatcgga ctgtaattcc aagcccgatc ctttcccgcg tgggttggaa ggcgacataa
50941 tgctagggga agaactcata tctgtaaggc agccaagatt ttatccctcaca cgtctctcaca
51001 gacctagcat gaggacttcc caagataatt tagcattgcca agtgcccagc atgggcccgg
51061 cacagtacaca tagtaggggc tctgataatg agccattatg gttttatgtg ttattattat
51121 attattattat ttattatttt tgagacggag tctgcgtctg tcactgaggg tttaggtgcat
51181 tgggtgcgac ttagctcact gcaagctccg cctcccaggt tcaagccatt tcttctgctc
51241 agcctccgga gtgagctggga ctacaggcgc ccgccaccac gccccctgtaa ttttttgat
51301 ttcttagtaga gacgggggtt caccgttggt tccatctgga gacccgtctg ttttccacc
51361 tcggcctccc aaagtctctg gatttYaggg gtgagccacc gaccccgccc gttattatga
51421 ttattgcat aaaggtacgt ttcaaaaata agatgagggg aagtgtgaca acgctctccc
51481 aataaaaacg aattctgcgc atgcagcgga cttcggcccc ctgtacggat aagRctgggg
51541 tccagagagg tagctgaact acccaagacc ggtcaagagg aaagtggaga tgaaggtggg
51601 gctctctgccc agtgtctcag tcaactcacat cccctctcgc ccttgaggta tcttttcaac
51661 agggcctcgg ggaactctgac coagtgtgta gcttggctcc tggagggggt caacctgggt
51721 ccgaaaaagca gagctgggga cgagatcctg agcagacctg cacacctgcg tgctctgctc
51841 tagccccggg ccagaggacc tcatctaccc acaggtgtgg acacgagca cggggttcta
51901 gaacttccca cctgcccgcc ttccacacca cctccagccc atccataagcc cgggcttctt
51961 cccacaaaata ccccccctgcc cctaggcgcc cactggagcc atctcacat tccagcgctcc
52021 cccctggctct ccttctcacc ccttctctt cctctcgagga ctccagcaag ccaagtctgct
52081 cccctggggg cctgcatttt gaagactgga cyccctcat cccacaggt gctctgacca
52141 catgaggcag agagaggaaa ggtctgcaca gccctcaccc acacaggg t cagcatcctg
52201 gcaactctcct cttaggtgtg tgggggtatg gaal-gagggg gaa-gagggg ttttaactct
52261 tagtgttagg cagaccatgg gcaggtgccc agaaagcaga ggtggacat agacaccccc
52321 ttcaataggg cagactcaca gggagaaaag aattccccct ccaatttctc ccaactYgag
52381 ctttaaggag gggccctggg ttgggtgtctg gccctcgctc taagtgaYgt gaactctcac
52441 tttaaacaga aaaggggct ctgttggag atattagtt cgagaggaa cgagggggg
52501 ctgaabaacg tcaactatg tctactatt caaacatcaa gaaacat at ttttaaaag
52561 ctttttaaat gaattaggt ttaaaagtg gcactggag cagctggag cagcagagc
52621 accggagact cctcaccaag taacctcac ccacga-tgt cagcagagc cagcagagc
52681 aggaatgtgt aacttgtct acctggggag cagcagagc ctccagagc tgtgtgagc
52741 acctccgctg acaaaagaaa tcggcagagg ctccagggac ctggccagc tggggctct
52801 tggcctcgcc agcccttgt tgacctccc ggcctcccac tggccctca gcccaccac
52861 tcaacaatgt catgaggtct agcaggtact tctgtacgt ctctctccc tggaaactgc
52921 ccaagagtg tctcaagccc agcagacct cgtgtgtgt gtcagtctcY cgcagactgc
52981 tgcgtgccc cgcctcgag ctgtggcgt gctccttag gaYgccagg ggcggctga
53041 ggctgtccag gctgtccagg gaggccccg caagggaggg aacacctc cagcgagcag
53101 gcaaacccac ccggaacaaa acctaccagg cgccagctg gccatctct cattttaaac
53161 tcaatagacc ctgaaccag ggtattatt cttttacag aggaaagcga ggtgagagg
53221 aagatggcag agacttgcca agctgtctca gacacagag gccccagaca ggaacctctg
53281 tgacttccag caccgtcgt tccccagccc ctgggtggaaa cctccagacc agcagggcc
53341 acagtggcct ccacaYgat gatcacacct gctgagact tgagagacc agcccgcat
53401 caccgtctgc cttatggagc caggggctgg cattcagca tggagaggtc ggaacctca
53461 gccgcagctg cctgaecctg gagtgggcag agccaccccc tgcaccact gcccgagct
53521 ggaatgggta gagccaacct ttgtccccg agggctgccc aggtgcacaa agctgggtg
53581 cccagggtgc atgcocagct gttatgactt gggaaacgga ggcocattct gggctgtgca
53641 tgttccccag tgtctcctg gcccctccc agaaacttct gctcaagct cactgaagc
53701 gagggccac atctctgggc caggtcccgg gcccccctcc caggtrgaag ctgacgcctc
53761 tggctgatt gaattgtgca gagaccaaga tgaacccctc tgtggcagct cgcctgagt
53821 cggcctctgg caggcactga cctcctgccc ccagggtgtg cccctggcca cttggggctg
53881 cMatcctggg gagtctgccc cggatatgct ccttgccct tgtgcagtg tgtctctgt
53941 gtgttctgtg tcttgggggc ctggagacca ccttgagct ggcctcaagt ccaagccctc
54001 cactgcgcgt ctgaaggcca cagcaagct ggcacagct actcagca tctgtggcag
54061 cccccagcca cactgcccac gcccceccaa gcccccagtg cctggagaag ccccttcccc
54121 tgaactggct tctgtgtcta agccaggtcc cagccccagg gtttccatg acttccccg
54181 cgaggtctct agagaactct ggcccgggag agccccagg ctgcaggtca tctgtgacat
54241 actgacccac agggcggtgg ccaggccccg agccccagc agctgagga cagcagctc
54301 cctgctgcac cactctctta caaagatgt tccagaacaa cacttttcc cccccagcc
54361 tctctcctgc tgcctcttcc aggcagggac cctctgtag aggtctgag ccttccggc
54421 gacagggccc ggaggcagaa caaggcccc cgaggcgcca cggccactcg gacagtgag
54481 cagcctggta ccaggccctc atcccgccac acgcttctc accagctgt gacagcaggc
54541 cactcgggtg acRrtggccat ttcgttttta cagatgggga gacctatggc atccccagct

54601 gggccgactct cggccgagaag ctctctctcct gcaccctgca gectggctca gacaagacag
54661 actgaacatg aagtttcagg cacaacacctg cagcccgctg gctgggcaaga gctgggacaga
54721 cagtgaggagt gcgaggcgga ggttgaggggc tgggaaagagt aagaggagta agtggggcca
54781 gggctggagca gctgtggggg ttccagtaacc ttcgggggctc ttgtgcccac ccccgactcc
54841 acagtctctgt gtccccctcgg cgcacacacac catgtcacaca cacacacatgt cacacacatga
54901 tacacacaca caagcacaca catgtcacaca cacaggacaca cacacatgca cccagggcctg
54961 tgtacacagc aaggacttttc caggggccact ccaggggacag ggggtggggcc gggcaggcctg
55021 cggggcagggc tgtggggagc cagagcagtc acttagctgc cttgtagtact cagggccaatc
55081 catgcccgttc ccagaggaag cgcgcagcaac ccaaggcctga cagggggggtt gaggaaatcgc
55141 cactctatttt ttgtacattt atttctttgt atttccaacy aacacagaca gctcaacttc
55201 aagaatcactt tgggaaggag agtgatggga atttccatag gaaaaactct cctccctggg
55261 gaacagggcgg ggcgttacc ttctgcccctg tcaaaactagg ggtcacacac gggcggtgac
55321 ctcgggcaaaa ccagaccaca gaggcactcg ttgaggtcac gtggtgccat ctggttgata
55381 gtatttgaaa actgggagat ttcatgtttt aaaaacttag aacatccagc ttctcttgag
55441 gaattaaagag atggggccac cccagggggtg gccKggctct gtgaattgag gggccacac
55501 tgggacaaa acacacaggg ctgggtcay caggccact ccttgccact gccaattgca
55561 ttctgggacac agaaggtatc acagtatgtc ctctggccc ctaaccgaca gtccccatgc
55621 aggggacccct ccagacatc ccgtaggag ggcattgggt cgcgtgggtg acagagctgg
55681 gctgggggac Rtcttgaggc tgcctctgcc gagcaggcag agtttttaac taactctgt
55741 gttgcaagc cattaaaac agcaaaaggt tttaagatt tagaattgta atttcaactg
55801 aaatggaatt attattgtgt ccatcattat ttaactcgtg ataatKtKgg ggtgaaattt
55861 gggaaattatg ttctagaSa tccctccacc caggacccag ccttagagag caggaggtgac
55921 gcccaaggcc aaagtgggtg agggagggaa gacaggggcc tggggccact gggcaggtag
55981 agcccaagtc tgcgtctggc cctgggaagc ctacgtgtgt ctaggtgctg ctaggacaga
56041 ggggagaagcc ctaagagcag ggggattcgc aatggcccta gacctggggc aacaggagct
56101 ctggcccgac cccctcctg tctcctctg tcccaacaag acccggcagt caggctgaag
56161 ccccgacttg tcaatcacga gtgccacctc agttttcccc acatataagt tgacctgac
56221 cctgtcaatc acagtgcca cctcagttt tcccacatgt gcttaacc ccagccatgag
56281 gagaacttga cccatgtctg gcccttgac acctgYtccc caatagagc cctgtatagt
56341 gggagctctga ccagtgggcc tgggttggtg ctgagctctc ttctagtga tcccttgaa
56401 gtaacctctg gcttgggcc tgcaccctg cactggtgct ctacctggtc taacagcccc
56461 tgytccactt gytccctcca aggcagaaga ggcctatttg actgctcccc cctccccct
56521 gggagactccc tgttagccc agccctgctg agctgtggtg actgggagct cacacagctc
56581 cctcgcaact cacccccctc ggttagtga tcaagcgacc catctctgaa ccttccccac
56641 agaggcagga ggtgacact gcccttgccc ttcttctctc ctgtgtgac ccttccccac
56701 aagaaggac tgggcccct gggggtgctg attgaccaag ccccaagaat cataccag
56761 tgggtcgag ctacctgaa gggaggtgtc tgtggcagga tcccatctt taataggta
56821 acacagact ttgagttcc tggcatccc tgtgtcagg tcccctttg gatggcagt
56881 gggagctctc cgtggccaaa Kaatgacact gtgggaagca ctcagaggt ctcagagcc
56941 attctgtttt agaactctca gaatggaaac cagtctgtc cagggaatc catcaaggga
57001 acaggagag gatggaatga gggaggcag cagcagtgcc atgtgaca gggcagctg
57061 accggTgggt caccgtgac acgggtgat accaccggag ttccagga ggggtgtgt
57121 gcaaaggcga caactgcag catctatggc acaccggag ctcagagga ctcggcagg
57181 cacagccctc cagcagtgca gtcacatcc atccaatccc taatccatc aacaacagga
57241 tggaggcgtg tcaacatcag catatcgtc taittcaaga tagggaaac gggcagcgc
57301 gcaggaaaaa tacatgactY ggaacagcca caccacagga gaggaggagt gggcagcgc
57361 cccggccctc gactccaag ttgtccact Rcccgagcc cttctgact tagaagaSag
57421 tggagggtga ttgctgRga tgagaattgt agaaacaagt atgattagac
57481 aaaaactggg actgactR caacctcggg acaactgtg gatgtgcgg gctctgaagg
57541 tgggtacagt gggcaggtgt gggatctgtg ggcaggtga ggggagcc agggctctg
57601 tgggcaggtg tggctctga aggcaggtgc agtgagcag gctggccct caggcgagt
57661 gcagtgagca ggtgtgggct ctgagggtg gtctagtg gacgtgtgg ccttgagg
57721 gggctcagtg ggcaggtgag gctctgtg cagagtgatg gctctgaag cagtgagct
57781 gggcagggct gggccctgag ggtgggttca gtgggcaagt gggggcctg agggcggtg
57841 cagtggcag cagcgggct tttgggcaag tttgggctc gaaggcagt gtaggaica
57901 ggcgtgggtc ctgagggcag gtgcagtg ggcgtggg cctgagggt ggttcaagt
57961 ggcagactg gccctgag ggggtgcag cgggtgcag tggcgagtg tggccctga ggggtgtg
58021 agtgggcag tttggccct gaggcggtg cagtgagga ggtgtggcc ctgaggcgc
58081 gtgcagtgg caggtgtgg gctagggcg ggtcagtg cagagtggt gctgaggca
58141 agtgagtg gcaagtggt gctctgagt caggcaggg ctttgagtg ctaagggtg
58201 ctgggtgttc cgtgtcccc gtgcgaagt cagctctatg ttctctctg gctgacct
58261 tcacctgga tgccttttt gctgctgcg cttgctggc caaatcttag cccctgaaa
58321 gaccagaag gaaggagcc cgtgcctcc caatggactc agagctctg aaagacaga
58381 gcatgtctg ccccatctg cctcctgcac agccacactg cgtgccacta ggggacgaca

58441 ggggacagggg ccggggcagg gatgcttaaa acctgaggac cacactgttc agtgacacgg
58501 ccgactgctg tccctctccc atgggatggc cgcctcccca tgcctctctt cctccaggtg
58561 tctgtgtgaaa tgcagctccc taaggagaaa ctttcccttat ttcactcccaa atgataaacta
58621 gccctgcccc gatttctccc gacagcacc cgtgtctgtg cttcctgtga cactccactga
58681 caggcaggtc ttgtctcatt tbtgaatttg cctgggacatt aactgacttt cccctctcaga
58741 tagagaaaccc tgggggatgg agaccatggg gtctgggtgt cctcagtgct cccagtgccc
58801 cacaggtlaac cgcagggctg tgaaggccct ggagcactgg cagagaggtg cctctgggaac
58861 agagggccca gcagagggcga gctggggccc tctggggagt tgggaactgg gatagaagga
58921 gaggcaggtg cagcagggccc ggaaggggcca cactgggggt ggggagagtg gaacagagagc
58981 aaaaagtgtg tggcagtgtg agaaagcarg gatggaaacaa aatccagcra caaattactt
59041 gcatatggca taaccaaagtc caaatgacca ggagaggtat aaacaagtggt ggcaaaagaaa
59101 cgtaccatgt Mattgccaac caaggcattt aaggagataa actgataaaa caaacaactc
59161 atggaaaaaa atacaaaaga taogaaactt tatgtcccta aagcttcaac ctggtgcaag
59221 gaaggagaga caaggcccca gtgataggag cactgtaata cattttctcc agacaaagctc
59281 gatcacagatc taggtaggga tagaaactat ttgaattaat agggaaaacct aaYgttcaaa
59341 cccaacagcc agaaacagctg acctrgtcaa gcacacagcg actgataaaa caaacaactc
59401 acataggagc acaactcaaa aactcaaaag ccgtgccacc agtcaccatt actgagacac
59461 attaggagac aactcaaaag ccgtgccacc agtcaccatt actgagacac attaggacac
59521 aactcaaaaa cgtgtgctcac agtcaccatt actgagacac attaggacac aactcaaaaa
59581 cgtgtgccac actgaccatt actgagacac actgagacac attaggacac ccgtgccacc
59641 aatcaccaat actgagacaa aatcagagat aacataaaaa cccataaaaa taactactgt
59701 aaattaaaaa caacRttctta ataggcctag actaaataga aaataacat tccactcatg
59761 gctcaagtag aaacactgaa gaaacaaata actcagaag aaaYtgaag agcagtcaaa
59821 gaaatgcttc caragaagag tttaacaaa Ycccaagga acagatactc ctcatctcac
59881 ataaaaattg gcaagatca agacttctgt aacataggaa taaccocgRa gaaagcagag
59941 tggaaaaaag agcttatcag ctataacacgc aagactccca agtatctaa gaaacagaa
60001 aactaatgaa tgttcaagac ctacacaaaa agtagtataa taatctctaa gRctccaaa
60061 aacacaaaaa aagacctcag taatgaagc tacatggaca taatcctgga cggaaagctg
60121 gtctcctccc aatatgaatc agaaatttga cataatgcc aatcaataat tgttaaaaa
60181 attgttataa ttattctccg tcaaaataac tccgtcgaa aatgtttaa atattctcY
60241 gtcataaata ctcogtcgaa taattgttga aaatatgtt cccgcaaat tccgcaaat
60301 aatctccatc aaataatgtt taaaacccaa caaagtgtatt ccaactttca cgggaagMc
60361 cacacacagc agaaaaataa Ygaaaaagaa tgacaatgag gaggcataat ttttcaaat
60421 taatgcagtg tattataaaa taatattaat gaaaaYtag gtagctcaat gctatgaag
60481 aaacagatca atggagctga ataaaaactc caaaaacagc cctgggtcag tggctcaag
60541 ctgttaattct agtactttgg gaggcagagg cgggtggact agctggaggt gggagtltga
60601 gaccagcctg accaataag agaaacctaa tctctactaa aatacaaaa ttagtgtggc
60661 atggtgtgtc atgctgttaa tcccagctac tccggagcct gaggcagagg aatccgttga
60721 accSgggagg cggagcctgt gttaaagctga gatcacgcca tctcactcca cctagacaa
60781 caaagatgaa actccgtctc aaaaaccaa aacaaaaaa aaaaacttcc aaaaacagat
60841 ccaagaaLat tgggggtttc caggttataaa tggcRgatg taacagagata ttaagttrg
60901 cctcttcttga gaatccattt aaaagtctcag taagaanaat taaaaaaat ttaaagtctac
60961 aagcctctYaa ggataggSag agcaggaggg gaagaaaYaa caagggaact gggggagct
61021 gaaaacagat gggcaagtgt taacagactt ggtacatcaa gaagactgaa tcccagctgt
61081 gccatgaaga cagccacatt ctacaaagtc tcccggagaa taagaagggg aactcatcagg
61141 tlacaggaga ggtcatagac aaagaatttg aatcagaagt actcagagt ctcaaaaaga
61201 gcatctgatg caggaaaaaa cgtggagaaa gcccttcagaa tcttgtagaa aatatgtgt
61261 ccaactggaa ttctgcatcc agacacattc cccatcaagt atgaagatgg aataagggca
61321 tttctagaca tctcagttct caaaagattt atgtgacaag cactctctc aagagactac
61381 tgaagaggtg gcaccatcaa aaacaaaaaa ggcaacaaa aaagagtga acacagatat
61441 agaaaacagg agcccaaac agatagagaa aacgggagcc caacacagaa gactagacct
61501 gggagggccc tggaggggtg cgagaagtga ccttgaggtg atgctctctg accagaYgtg
61561 gagagcagat gagccatact gagcagcgtg acccacgggg tgaactgtct tccaacagat
61621 cccactctcca tcaggccatc tgtagcccg ggcacaaatg cccagagagag cctgtgctag
61681 gttcttggca gtacttggct tggccatgtg cgtatgaggg ggcctcccc cctgtctga
61741 ggacgcacaa caltccacaga agtttccagc tttcccccca ccgagagagct gagggagccc
61801 gtggcgagct tagaagcagg acgtacaggt cagccccctc cctctgacgt Rctttaagct
61861 gcaggttccag cgcgtgtgtc ccaactgggc cctaccaggt gctctgactgt gctgagccca
61921 agactctctc gagaccttgc tccgaacttc cccaggaggg gcttctcag tctccggaga
61981 agcagaagct agaccacagc ctggagagct atctgtcac cttccggagc cttccatcag
62041 cggggggcag cgcgccctct ccttacactc ccagcgtgtg ccagctctcc Rctttaat
62101 taatcgtctt tgatattcat ttgcatgtaa ttataaatga tgtctcagt agcaatacag
62161 tgaataaaaat tagtaaatga taccaggtag aaacggtgct ggaagacggc gaggtrgag
62221 caggagggga caacctggca gtctggcact actgagggga tcatgaggaa gagccactag

62281 cggctccatct cctaaagata caggatgtct atcggatagc ccagtgccct ccttggactg
62341 acccttggcc catgggcat tgttcggccc catggactgg cttgcagatg ggcgaacttg
62401 cacaggtctg taaggaggca atggggagct gcccttgcga gcatttgcctg ggtggagtga
62461 ggggagctgg aaacaacacg ggggtcatca gtgggtgaga ggaagaagcaa atgtgggtgg
62521 ggcatccccc aaagtgtcac caaggttaga agtttgatat ctacatagat ggaatggatct
62581 taaaaacaga gtgccaaagt caaaaggtaa gaaaacagaat gagacacaat tcagagcgac
62641 ttacttcaat tatggatata cacagcgcaag acagcaacac rcattgtgcga aaaaatgaca
62701 ctagggaag gatacgggtt gatcactgta aatggttgct cactgcagga gaaaatcaga
62761 aggaagagag caggctgggg agataaatgg cttttgtttc ttcttttgtt ttaatgtagt
62821 taggacagc gggttagaggg aggaaaaggc caaatttga cctcatcgct taagcaaaaa
62881 cagatrtcaa gtggatgaga tgatcaaaag gaagtcatgg aagtgtcaga agaaaatgta
62941 gckaggaccc tttttatcat ctgtctgttg ggcgggctt aggttcccaag cctaggggctc
63001 tttcccgagc ctctccttag ctgcccacac ccttgtgagc agtttcccaag cttgtggagc
63061 tctgtcgctt gaatgcaagg gcagatggca cacttgttag gggatgggtg gaggggggatc
63121 ggggcacatc Wctaggggat ggcgMaaggg ggggatgggc acactcacta ggggatggcc
63181 tgagggcagc caggcacact cactaggtag gggatggctc gagggcagac gcacatgca
63241 ctaggtaggg gatggcctga gggcgacagg cacactcact gaggcaactg cacacagaca
63301 gagggtgtgt gctactgctt caacctgcc aacctgtgga tgggaacaa tcattctctt
63361 tccagtgctt aggagtcctt aaaaactgaa tctaatgac cggccYgac agccagaaat
63421 gccacaStgc ccagaagagg agagggaggc tcaaggaaat agactgaggtg catgatgttg
63481 gcaaatggcc agggcagagt gccagccagc tctctgcaca caagcatctg caggggctcg
63541 agggggccacc accRggctca ggaggaccaa agcctggagc accagcagac tctctccaa
63601 catgagagc cccacccagg ggagttcaac gggaaacagg tcatccctc cat tctctac
63661 Yaaccaagag agagtgcctc tgtggctgtg aatcaggaga cagcagaaa ccctgtctgg aagcagcgac
63721 atagtgttat gtggaaagt aatcaggaga agcctcagcc ccctggggggc catcagcaga
63781 gctcttttga gggtctggga ggtgactct attaaagat gctcatagg atattacat cccagccctt
63841 gagcccactt taaactctta ttaactgact gaaatlaatt ttgtcaactt caacagaca agctatcca
63901 atgcaatttga ggttgatcgg ggttgatcgg ggttgatcgg aggggtggga ggggctcctg gggcgggctt
64021 ggcatcagg agcgcatcag gaaccgctgt ggcgaatctc agggcagctc ctgacagctc
64081 tctaaactcc ctYggagtgg ggtgcccgcg ggcagatcgg agcgtccagc cccagcctcg
64141 tcagctcctc ccRgtctca cgtgagaaaa ggaagacata ggcagagaga tccagctccg
64201 gggcagctcct cagaacatac cccgcacagc agagacaccc aggacagccc tccagctccg
64261 acacagaaag tcttgccctg cccagcacct gctggccat tacaactttt ctgacagtaa accacccctc
64321 aaacagaggt ggcctgccc tctggccat gccaatccca tctgaccccc gctccaggag
64381 cctatttgcga ccccatatg gctgatgctg cctgatgctg tgttgttggg tgcacagaa
64441 tgcccaggag gctctgttag gccactggaa atgtgactct ggtgRctgta agtaactcct
64501 gactcccgat ctttccctct tgcctggaaa ggcagcccc tggactcctc aattctcag
64561 tcccocagg gacagcagag gctcatctga cccaagag agctctgtgc tctcacaac
64621 aggttcccaat atccagatgt gcttgaagtt gatgaccccc tggactcctc aattctcag
64681 gctaatctat ttatgtctct gctcttgcga accacagag tctgtgtgct tgcagccaa
64741 acacagacag caaatgtgc tggcctgaga gagacacat ggtttgtct tgcagccaa
64801 gctctcttga acgattgagg ctctcagacca ggcactctc tctctccctg tgcgtcaac
64861 tcccagtgct gggcagagag aaggtctcga gtatttaaag tctatgacca gaaaaggggc
64921 caggctctgg tccggctccc agatcgtgt ccttgggaaa tgggtccagc cagagaggtg
64981 aggtctctga ggttagagccc agccagcagc ggcctcctg agtgaccagc agctgtgga
65041 gccctctga ggtgtcctgc agcagggtgt ggcctctgt tctacagccc tctcagaggt
65101 tcccaaacac aacttccccc tcttcaaccc agggctgggg cctctctgtct acagccccc
65161 cagaggtctc aaacacaac tccctctcg tcccccagc acccgccgaa taccagagg
65221 tctctctgac ctgcccctgt ccatcttcca gcggcYctg cagcccactc accctcagg
65281 attcaggggg gctctggggc ctctcagagg tggggcggtt ggaacagacc acccaagcct
65341 gccctctgct agcatctgct taagccaggg aggttcccc gaactacca cagccccagc
65401 ggctccaaag ctctccctgt ctgggcagtt ggcctcagga agggaggtg gcttgggtta
65461 tgtcactggg gccccgcagc acgcggcagc caagtgtgtg gggccgatgt agagggccg
65521 aggtcttttga tgaataatat taactgtgaa ctccagaca aaacgaagat ccccaggg
65581 cccaaggaga gggagaattc ccagggggtg tgttctgaac gcagaRagg tggggggcgg
65641 caagtggggt gctgacagcc agtgccacgg gacctatgrg gccggagggc tggccagtg
65701 ctacagcagat gacaggcgga aatccctgt caaccagtYg gtacatgctc cagagccctg
65761 gatttggaca tcatctggaa gcttggcatt gctccccga agcctcagc taactcaac
65821 tcaagaagctc caggggctcg tgctctctct gccgcagct gcaactccgg ctgcagctg
65881 agctgaggtc tccaactagg gagtggggcc cgcagcagc caggttcatt cactcaagc
65941 aaactcgtct tcccctccc tcaaaggcag ctccctctc actctcagc gttcttggaa
66001 cctacccctc cagctctat gccctctgta ctgcacacg tcttgactcc agcagatct
66061 gcgcctcccc catctctgca ctcttcaaat ceaagtcca aggcatttct ttacttgcga

66121 acctcagttt cccrtgtrgt ggccttcaeg tccctttagt ttgagccctg cgttcagtga
66181 ggggtggcgt gggccctgca ggggtgaggt cctgggtccc tgggaagagc ttgacagcaa
66241 cacaagcggg aggtctcaga tgcctgccga cctgtgcgaa atgagccctc ctctcaccgc
66301 gccatgagcg acgtgcaggg actcagcacc tgcctctgct cctctcctct cctcagcccc
66361 tctaggtcag gagtccaggc acacagcagt gYtggggtgc tgacagcaag ccagggctct
66421 tctagcttga ttgtgggacc tggggcagcg ctgcacggtg ggcacacagt tgcacacaa
66481 ctgccttctc agggaggaaga gtctctcaga tgcccaggca ccagggactt gcaacacctc
66541 ctgtttcttc ttcatggtag ccggtggagt gggcacggga cctctcttac gcaaatgagg
66601 gaacaaacccc tgggcccagg caacaggcac ccttgcgtga ccccaagatc tggctctgct
66661 ccccgcaccc agagcttcca atggcccatg ttgggtgatg ccactgtgac accagccaaR
66721 agggagctga aagtaccagg gccctggccc agctggaggc cttgctgcca tccaccacca
66781 tgacccttgg gcaacaaagac ccccactttc cagtctgggc accaagcaca gtcatgtct
66841 gaccaggacc tgctcaagga gcaggagtga cagaggacaa ccccggggac ccccagcct
66901 ctcccccatg ctgggggtgg aaagggtgggt aggcctgggg cctcttgatg ctcccatgcc
66961 ctgatgtccc atctccagg acgctcccat gccctctgcg cctcaggagg ctccatgcc
67021 ccccatgtgg cggggcagca gccaggctcc ctggatctgc ccagagaaga gcccoccgca
67081 gagggtgatgc aatgctcacc acgccccctc cagctcaaga ccttggactt tgcgagcccc
67141 tcaagcctct ctcccagtga gtgtcaggga cctcccttcc cctcagctgc acccagcct
67201 caaaacacat gtctctcttc cctctgtctc agccccctat ggggctgagg agcaggctg
67261 cccaaacctt gtctctgttc tctcctgtcc ttcacgcctg gctgtgggat ctctctcctg
67321 acctgctctt cccacgggta aatcccaagg tgcagactca gagacgacac cgagaggcccc
67381 ccagtgagcg ccgcaactctg agcccagagg agcccacalc caaccttgca gggagcgggg
67441 ttctctcact ttgtcttct gtgtgaactc cgtgggtcca gaaggctgag tttgggggtg
67501 gtgacaaatg gccctgggtg accccacaca cctctgccc atgagtctc ctgcctcaga
67561 atatacccag tgacgttgag gagggaaggg gctcctdata ctccccacac agcacccaga
67621 ccaaggtggt ccaagaggtt aatttgatac gagaaggatga gaaatcagaa gaaatcagaa
67681 atcagagtga gtgctgagtc ctgaaaaatt aatccccagg cgagcctcac aggggcttct
67741 ctccccacg ggggcagggt gccaggatgc actacaagcc cagatccccag gcccacctg
67801 agaacaggga ggggaggagg gccaccctgc cctgcagaga ggatggcaaa ggccctgctt
67861 gggggccccc agctgcctc aaacttgcaa gggctcagaa gactggacca gggcgtggcc
67921 gggcgagaaa latagaaatc atcctaattc taacttaaaa ggctgggggt gctgctgctg
67981 gcttaaaagc tgggtgtgacc ctctcaggta gcagccagcg cgaggcctgt gctgctgctg
68041 gatgggctgg gtgggtggcc ctctggcctg tgctcccaac cagctgccca ctgctgtgta
68101 cctggccctg agaggtgagc ttgctggcat cagagcatg agcagccctc cgaggggcca
68161 gccaaagccc atgcttctcc cagggcagga gcacctccc cagagcaccc cagagctgct
68221 aacctccagc ccaggcaacc cctcccccct cctccaggto cccagctgact agtctgctg
68281 ccaacttctg tctactctgc ctgcccagg ccttccattt tcatctattt gtctatgtt
68341 ctccaacagt tccaactcac catttgccag cctcagtgat actcctcaga ctatgcata
68401 ctctctctct tcaagratte tgtaactcac cacttgccag ctgagtgatc tgcatacaca
68461 tctgggggtg caccatcaca ggcctggcctg ctaacacaca cactcactca caaaatgact
68521 cacaacatgt ccagcagtrca catacacata cactcactca cactcactca cactcactca
68581 cacaacagaa cgcactcaca tacaacatg cactcactca cactcactca cactcactca
68701 cgaatgcact cagactcaca caaacacgca cactcactca cactcactca cactcactca
68761 gtacatata cactatgact cactcactca cactcactca cactcactca cactcactca
68821 cactcactca caaatggact cactcactca cactcactca cactcactca cactcactca
68881 tgcactcaca ctcacagaa tgcactcaca cactcactca cactcactca cactcactca
68941 ctaocgctg aatgcactgc cactactaca tgcactcaca cactcactca cactcactca
69001 cactcactga cgcactcaca cactcactca cactcactca cactcactca cactcactca
69061 cacaacagca catgcataca tcatcactat gtaggtgat gtgagtgcat gtaggtgat
69121 gtaggtctat gtaggtgat gtaggtgat gtaggtgat gtaggtgat gtaggtgat
69181 gcacacacta gcacacacac aaactgtgac atgacacaca aggtcttaca gtaggtctat
69241 ggaagactga ggtggccctt gacccctctc taagtccca ggtaccaca tggggctgtg
69301 caaaaatgaa aataaacgg gccaaaaagc cctatkgat aaarogaaa cttagttatc
69361 ctatgatctg ttctgagcag gacaggtag acagggtag acagggtag acagggtag
69421 cggggctggt gctctgatat ccttggagg agcagtgagg cgtgagctg acatcggtt
69481 cctccctggt cgtcattacc agggagagca tgaaggcgag gtaggttag cgtgaggt
69541 ggttctggtg cgtctactgc agcctatcgg tgggtgcagg ctaccacagt ctctcctct
69601 ctgtcttggg agctgtgccc atggctccc gcacaYctg atactcagat gctcctct
69661 gacaacacgc tgccggccgc aggcctcaca ccaacagaa tctactcagg gctcctct
69721 cctcgaactg gtaggtccct gctaggtcgg cgtgtgttcc ctagtcccc ctctctgac
69781 cttagcatat ttgttcaca tgttgaacag agagaaaagg tctgcatctt ctagtctgac
69841 tataactctc caatttccct tggagcatgc ggcctctctg ggcctctctt tctgtattct
69901 ctcttcactg agaaatattt tgcctccctg agaaaagcca tgtgggcagg gggatgagag

69961 ggcacactgc cccagcccca aggtggccag tggtagactag gagggtgcag gtccatgttg
70021 ctgggtctta tttttgtaag gaagtggagaa atctggatgt ccatgtgaag tctctgatct
70081 ctaaaatcgt gcaaaaaagca atttaagaatt aatgtagggt cgtgtgctgt tgtgaaatgt
70141 accaaagctg ggtcccagg tctgtctgcc acctcagctc cagcccagca ccaaagctgt
70201 gtctcagggt cgtgtctgcca cctcgggtcc acgcccgtgag caaccgtttc cctgtatcct
70261 cctccgtctg cctcggagag gttttgagaa cagtaggttc tggccaacca gagtgaagct
70321 acctgcctct gctccacagt ggcctgaagct tcccggggcg cagttaacgc cccacttgtt
70381 tctccacalc cctttccacac ctgtaaggac agctcctgat ggatgtgttt tabggtggaa
70441 tctctacgca ggggctctgc gaacactccc tggttatcca caccggcaact aaactcagc
70501 acgcaccagg cctcctctcc ctctcatgga cctcaattc ctggaacac agcagaggc
70561 tcaccacat catgtgcgtc tcagacaaag gactaattc cttaatacat aaagcgctct
70621 tacaatatcaa caRgaacaa tccagcaatc caaaagggaa atatggaagc ggtattgata
70681 caaaagtgt gggtagagaa gggcatgct cctttaaagt atatggaagc aggggaagtgt
70741 atgggtccctg gctaaagggt ccacccctcac ggacctatgt accatgtccc catcttgttc
70801 ttgtgcccaa atgtcacatt tcccagaacc acctggtccc accatgtccc catcttgttc
70861 cagagaccct aacaggcaga ccaggcgggc cagagaggaa cgcacaaagc aggagagcga
70921 tgaggggagg aacacacagg cggctggagc tcgagaggaa cgcacaaagc agcaccagca
70981 cgctcgaggc caccagccgg cagaagcaga acgacacgaa gtttggccgg gtagctcaga
71041 ggagagccca gggtgcttag cggcccgagt ccaggggaaa ccccttgcac cccatctccc
71101 ttctgggttc ccccatctgc tgagagctac tccactcaa taacactctg catctatctt
71161 ccaaagccac gtgtgactgt attcttccag gacatcaagg aagaaaccga gatcacagaa
71221 gctcctgtgc attgcgacaa ggttagaggggt ctactctgag ctaactgagc agcccgctga
71281 cagacagcta aagtaaaaga gcatatcgag cacagcccca ctgggggttc aggggtctga
71341 aacatttcca cctagacact cgtgtgggggt cggagccccc cagctgccc ccttgtatct
71401 tcccacagag gtttgagcag aggggacactg aagaaggaaag ccacaccccc atagacagcc
71461 ctgcaggagg gacaagagaa cttctctcat tccaataga ggggacaatt tcgacagaa
71521 aaaagcagc agcttttctt ttctcttata tctcttctt tcttttctt ccttcttctt
71581 ttcttcttct ccttttctt ttcttcttct tcttcttct tcttcttct tttttttttt
71641 gagacagagt cctgtctacc caggctggag tacagtgccg tgactcagc tcaactgcaac
71701 caccgctctc cggatccaag cgtcagcct cctcagcct tccaagcagc tgggattaca
71761 gttgctcgtc atcaatcctg gtaatttttt tttttttttt gtaatttctg tagagacagg
71821 atttaccat gttagccagg ctggtttcca actcctgacc cagggtgat calccacctt
71881 gcttcccac agtgtgtgta ttacgggatt gagccacgca gcagcttttc atgtgtgag
71941 aagacgctcg acctcacctc taattaaaga aacaaactcg atcagattgt ccaaataaaa
72001 atgtacggat cagggggtaca tgctgagaga aacggatcca gccaccatc ctagggttgc
72061 ggtcatcctt atcaaatgtg aaaaatgcaa aaccttttga atccaactt ctagtattca
72121 aatttaactc tacaattgtc cctatacag tatccaagga ataatgtga aggatatacca
72181 sggtcgtatt ataagaaacR gtgaaagaa gtaacgtccc catcagtcag ggtctggtta
72241 agcaagtgat tcacaggatg gaatgcctga gcaggagaa tgggcagctc tggagggctt
72301 gctgtggagc agctccaaga catgttccct gataatcga aatgcagggt agcacacggt
72361 gtagctacaa gtgtgcagaa aaaagtcaaa caccagtatc gaacctgtac gactgtggg
72421 aaatccctgt agccagagat tctaactcta ggatttatat atgtgtgtgt acaNattgttc
72481 atatatatat acttgtgtgt gcacaggaa tcttgaaagg acacagaga aaggtgtgca
72541 gtggctgccc tgcaagaagg agctgaattc caggacagaa aatgggaggg ggaatgagac
72601 agcaatgagc atttcatatg tgcgcaatcc ttagtccagc caaaaagcga ttcagcgca
72661 tgcaaggagag gccaggactc tcccagcagc tcccacagg gtacagggca gaggggagcc
72721 ccagctggc ctgctccctg agcactctac acttggcagg tgaagggggc caccctgaca
72781 gtcaaggccc agctgtgatc tggggtggaa ctggtctggc acctgtctg ttggctttga
72841 tctgggtgct ggcacgcagc tatctgggtg gctgcaggc cccacgtgtg ggccctttga
72901 gtgatgatga gctgtctggg ctgggaccac agagtgaaga ggaaccagat cccctcaacc
72961 ctgactgccc tcyaggactg cagggcccaag acacagaga ctgctgtcc aaggaaggag gggccaggga
73021 agggctgaga ggcaggcaga gggggaaagg tgttctctac cagccgccac aggacacga
73081 gtcagagggg cctgggagtg agggaaagccc ctgctgtccc aaggaaggag gggccaggga
73141 ggagagctgt ccagagccggg gaggaggcca gaagtgcaga gaagtgcag ggggagggc
73201 agcccttgct tggggctgga gaggagcca ggaaggagg tggaaagcag gggagaggc
73261 gtgtgcgcaa agactgagag ggggaagtac agcttgggtg cagaagaatgt gggcaggag
73321 gactccgggt ttctgacctg ggcaaccggt caccctggga agcttccc ccaactggac
73381 ctggggaggc gccgcgtggg agggaggaaa ggccgtgtgg ggctgttgg gctgagattt
73441 ctgtggggga acaaggacag acaccaYga gacactcag gacagtgtgt ctgagcaca
73501 gagagagggt gggccggagg cccaggtgtt gggcgccctc caggaatgaat cagctcaga
73561 gccaccaggt gccgcacaca catcagtgct ctgactga aaagtaccat aagtagcgt
73621 gccacagccc caaagttaaga cacatagaaa acctgagccc cagcaggtg acaggaagcg
73681 gccagtgcca cggacagggc tgttctcgac ccaactcag caccagga ccaactctgg
73741 caccagcctg ttttctgttc cttccagcc agcagggagc aattgggggt ggaacttccc

```

73801 cagcctgaag acctccgggg cctctgctcc tccctgctcc acctgccata gcccaaaacc
73861 agtgtgggtg tgaagcggaa ggaagcttYg cagcaaggcc cagctcttat gccggccact
73921 gccttgctcc ctgectgatg cagtctctcg gcagcaaggc aggtgcctct agactggagg
73981 gtgcccagag cccacagaacc caccatgcgc ggcctgggcg acgtctgagct ccccttaagcg
74041 aggcagacaac cttaaccaaag ccagcggaca tccactgcct gcggagcatc tggggaaagcg
74101 tccgttgggg cctccagacc ggggcggtgc cttcaccacg cctctccatc ccagcatcca
74161 caactctgga gctaggaagc ccggtgclcc ttcaccaggg agggagcgga agctctgcca
74221 gtcctgactg aagctclgtg lgtgcgylca cctgagaaca cctcagtcag tctggcttga
74281 tccctgactg ctcccccggc ctgtgcttga gaggtgacag ccagcagcgtg agcaagtgcc
74341 ctgggtctct gtcccagcaa agctaactcc agctgtgcac cctcgtcagg caccctccagc
74401 tgtgcacctc ggtcaggcca cctctctccc gtccagacct cagtgctctg atcagtgctc
74461 acaagtaaga cacatagaaa acctgaggcc cagagaagcc aaggaagtcc ttacaggctt
74521 ctgtcagaga atcgaggaaa aggaggggag gccagagaat ctgggcttaa gtgacttggc
74581 cgtgtcagaa acacaggcag agtcacaaac acgagaccct gctgccactc cccgacggag
74641 aggttaattcc acaaacacga gaccccgctg cccactccac aaacatagaga ccccgctccc
74701 cactcccgac ggagaggtaa ttccacaaac acgagaccct gctgccactc ccaaaaaacc
74761 gagaccctcc tgcccactcc cgacggagag gtaattccac aaacacagaga cctctgtccc
74821 cactcccaaa acatgagacc ccgctgccca tcccgcaggg aagggttaatt ccaggctcgt
74881 ggcactccagc ctctcccgcg ccccaggccc ccagctgccca ccaagaaaaa ccttttgtag
74941 ggggtctcgag ccactggggg gcagagggtac atcccccaag gccacggccc gcagtagacc
75001 cttctctctcg agccattctt gaaatcacct ccaagaaaaa agaaatgact cagcttccaa
75061 agatatcccc aggcaggaaa tgggcgcgag cgacgatgaa tggtaattct cgttaattgt
75121 ccaagttccag cgcacatcat cctctctaag taacctatg gaaatcgtaa tggcgctga
75181 aatggctcat ttccaaact cccctttctc gggggtctcc agggcgctag cccctgtcag
75241 ctggcggagc aggcctctct tgccaagggt taacctctct ggttctaat tttaatttgt
75301 acctttttct tgcacatcct gaccccgctac atcccctgct cttggcacac tgcaccaaa
75361 caccatagaa ttactagact ataagcctct catggaaatc tggttttgac ttgtttagaa
75421 aacactataa tcaaatcccc gagagtcagg aggagtgggc ttggctctgc ccaaaaaa
75481 tatcaccgga ttaagtgat ttagcagagg ggtcccgcg gggagggggc agggcaggct
75541 gggcagggct gggcgtgag gggcgaggca ggggtgccctc gggggcaagg ggggtggggga
75601 tgcgcagagg cagaggatgg gcaattgtgg attaccagaa agtgcggtct ggcactcccc
75661 gggcctcagg gcttctgag tgccactcgt cctcagggca ctccagggcc atttgatggc
75721 agatgctcag agtgcagag ggtgtgcagg tctggtggtc acttatgccc tccagggcct
75781 gcttctgccc atgccagggc aagaccacgc ctcaaaacac catcaaatc atacgttga
75841 cgttatgcgc gggagggctg tgcacacagg cacacagaga cggtgagact tggggtaact
75901 tgggtccagt taactgccc ttgggtctca cttctggag cgtgatttg gctgcaccac
75961 ctgcactcgr gtgtgtctgc gtcctccga acttctccca ttctcagcat gtacctaaaa
76021 cataccacca ttcattgact gtcaggcact cttctatttt gtaaaacat agggttata
76081 aggcctgcgc agtggtctcag gcttctaate ccaReacttt gggaggccga gggcagggga
76141 ttgctcgagg tcaagggttc gagaccaggc tggccaacat ggtgaaaccc caactcagc
76201 aaaaatacaa aaatlagccg ggcattggtg tgggcgctg taattgcagc tactcactga
76261 cctgaggcag gagaatagct tgaccgcggg agggcgaggt tgcagtagc cgcagatcgtg
76321 gactgcactc cagcctggcg cRacagagcg aaactcagtc aaacaaaaa aagaccggcg
76381 gggctaccaa anagtagagg ggaccacagg ctgtattgca catccacatc agtgaccgtg
76441 taattgtaYg cagaccacct ccccRactgc acacgcatc tgaataaagc atggaccaga
76501 aggaacaaata cggcacctgt ccactcacat gaggatctgt aatcaaacg atgaagaagc
76561 atgacacact agagcgtgcc agagctgggg gagcgggaaa tggggagtgt cgttcaatg
76621 ggtgtaaagt tttagtgatg ctggacaggt aagtaactga ggtctgaggt acaagtggtg
76681 cctgttaact gacatagtgg cactgtgcac ctgtgagct ttcagactac cctgtctcac
76741 gttacagtgt ctttccacaa aagcaaaacaa caacagcaac agaaatgaag kgacaggaag
76801 aaacgctggg aagaaacgag tgtctaaccg tctgtgtgc cctgtgtgc cgggtgtctg
76861 cacaatacca cacttgctga actgtaaggg gccggtgacc tctacagctg cctaaccgtg
76921 cagactcaca cttgcagaga agagatgag aattctgata ataggacaaa cactcactta
76981 acaagttaatt aaaggacaaa ttggtgcata taacaagtgc ttccctgtct cgcagctgac
77041 atggatcaat gtccaaaatt ttatccaaa ttgttataaa tacaactgag agtclctctg
77101 gtgcagtga taagtgaagt agcagaggca gtggctcacc cctgcagctc cagcactta
77161 ggaagcgga ggggtaggat ccttgagta cacttgagta aggtgcagb aggtcagctg
77221 tgtgccactg cactccagct ggggccacag agcgagaccc gtctcttagg aaaaaaaaaa
77281 aaaaaaaga aaaaagagg aaaaatagca aaattagcca atatccaga agtgtaggag
77341 agaacaccga tcaacttttg ccaaaaatga gaactccaca gttgcctcgc cattctctc
77401 tttaaaaaag gtcaatttta cattatgaaa tagagtgcac ctttttaag cattactcga
77461 gctccaggaa gatattccac gtttatcctg gacttttgtg cccctctcag actgtgtttt
77521 gattaaaaa attccaaact cagtttgcta agccagtgca cctgtgtgcg cgtgttact
77581 tgcaccaatg ttttatcaaa tctccgacta tgcccaagtg cggcgccatg gaagccacac

```

77641 gggatccctg acagacagaa acatccatac tgcctttaag gtggctcagg cactgttcca
77701 agcacctcgt ctatagrtgc ccactgatcc ttcgcagcag ctgtgtgagt tagctgtgat
77761 gcttatcttc atctctgcaga tggggaaact gagaaaagct atcccaattt cactactgctc
77821 ggaagtgtca gaaccaggat atgaacctgt gcatggctac caagagtctg tgctcttgaa
77881 cgctgtgccta ggccactcct ggccttgatc ttacatttta actctgacat gtccaagaaa
77941 tcaataaag ttatttgagt gacgtatccc ctgtctcaic tgtaaaagtc ggtgacaagt
78001 gctgcaggtc ctgcccactca gacgtatccc acggtgtctc gYatcaggat gaggatgagt
78061 agatcggttaa agtgccttgt agactgtgat gatcatattt tttcactgctc
78121 agcaacaatc cttttttttt ttaataaaagg gctgttgaac tgcattcccc aattcttgat
78181 acattctctc ctgtttctct tggtraaagt tcaagaattc tcttagggtt caaatgttta
78241 aatttccaaat cctttggctc atctgactaa ttacaagcat aaaaatattggc atttcccaac
78301 agtctggaga atgggaagcac atgaaattat gcacacacat gagcatctgt gtcgggaggc
78361 ggggaaggcgt tcaccccttca aactcaagtg aaaaaggttt gccgtagtta aaggctgcgt
78421 acaaatcaact actgttctga gaaatcacca ctgtttactt ggagaatcgt ttttcaagca
78481 atatatcaag aaaagtaaat atcaagataa atttgacag ttgcagataa ctccacattt
78541 tggaaataaaa aactctgtgt tctgtgtaca agtgtgcaaa acacagcaga ctcctcccc
78601 tttccaatcc tgcttaataa ctgactgaac atcatcagtg ccatctccca ctactctgt
78661 agtgtatctt gtgtgtatct tgtgtatcat attatctgct tttgttaaat ctactctgt
78721 taaaatgtgt cacattaaaa aacaaaaaca gcccgggcat ggtgtgtcac gctgttaact
78781 ccagcacttt gaaaggctga ggttggggga cctctaaaaa atacaaaaat agtttcaagc
78841 cagctctgccc aactgttaat cccagctact cgtctgagtc gaggcaggag aatcgtctga
78901 tgttggcagg accgtgtatc gtgagaggag attgtaccat tgcactccag cctgagtgat
78961 accggagggg agaggttgcc tccgtttcaa acaaaaaaaa aacaaacaaa cctcaagcgc
79021 agagtgaagc tccgtttcaa acaaaaaaaa acactgaagc cctcaagcgc ctagcagccc
79081 aaaaaaacag aacacacacac tacactgact acactgaagc cctcaagcgc ctagcagccc
79141 tcttccaggt ccccaaggac caggcaccaa ggtctggggg tggccacaga tccatctcc
79201 cctttctcag acgacagacc cagtgttatt tectggaaat cctggaaacc tgcattgaca
79261 cagccctccc caagccaaac atctctcccc tcacagagtg tgggaattga gcgcagagag
79321 ccagaagagg aaaaggtagt ggtgtcacta aacacttggt ccaaatggtt aagaccctcc
79381 tgtgtccagg accccgactg ccaggccctt ccaagccctg actccccgtt gactctgag
79441 gtttctaat gtctctccag taactcttcc tggctcaggg cagctggact tcatggaagc
79501 gttctccagc aagagaacct cctctgccac agggccactc ccagctgcat gcagacaagt
79561 tcaccagccc ctgaggctta accctgtctc tgtgggggat tctctgtgct ctgcctgata
79621 tcaccagggt gggcagctct cagggtgact ctccagctct catgtggaga ggctcacaca
79681 tgcattctagc agagaagtgg aggtgcacc cccacagaaa gtggggccca gtcccacgtg
79741 cccctggcag tgctcatcca ctaggactgg gcttccgtgt gcagggtctc accctgtctc
79801 tctctctctg ctgtcccaac cctgtgctt tgacctccat gacctggagc ctgctgacag
79861 gtgccccccc acccaactcc agcctgcacc tccccaggct ccagaccgtt cccccgtg
79921 cctgcaactg cccctggggt gtccacaagg acctctatgc caacaagtcc caaaacacac
79981 tgcaccttt cctcaccca acccatcagc catgtgtgct gttgccatct cactctccca
80041 tccaaggagt caccgatcc tggcaccttc cccatctccc cgttccatgt gactctccca
80101 cctccctctc cctccgtgca ctgagctcct cccagactcgt gccccatcgt cactgtctcc
80161 ctgcccctca ctcccaagc ctgacagacc cgtcggggag tgcaccaatc aaaaactctc
80221 gtacctccct ggcctgact ggcctgact gctcccaagg accctgacat ctagctaggg
80281 cctctgtagt ggcctgact ggcctgact gctcccaagg accctgacat ctagctaggg
80341 cgccctcaca gcatgggtc cgagaaaaaa ctgtctacac tcccagctcg ttagggccgc
80401 cctttctctg cagactctta ttcagttctc tcccaagacc tgbtgcctct tcttttcaac
80461 aaagagatca agttaactc tctctagac gtttcaagt cccctaccca gacctgagc
80521 cactgagca cgagaagtat aagctaccca cctctgggta cctgcttgg ctgagagagc
80581 tccagggtga aaatcatgca tcagattcat aggaatagca atgagcagca atcagagctc
80641 ctctctgaag gtttggcaca atggagtaa ctaggggttg tggagacaga ctgacgYtt
80701 atctctatgc ctacagaatg catgaaatgc ccaagctatg tctccaatg tccagctgct
80761 cacaccaatg gagccagRtg ccagctatg tgtgtcgaag actctcagc aactcactg
80821 catgaaaatc agagggaacc caaatcgaat gagcagagg gctgcacaga aactcactg
80881 gccctttctg ctccagctct aggggttgg ctgtccacga actctcagc agaggagct
80941 tgttccacc tagtcccaag catctgaaga gggccaaagc aagtgctgt aggcacagc
81001 ttgcccctga gaggaagggc agagaggaa gaggagggaa gcggggtgga cccactctgc
81061 tcagaccaaa ggcagaatcc agtgctgat gtgtccataa gggctctc tttgtctggg
81121 ccccactgaa ccatgtgagc cctgcacaga cctctgcta gctctggg gctctccg
81181 actgagatgt cctcaacagt gcccccacc ctagggcatg ggttgccctc tgcactagg
81241 caagtgcctt cccagacacc ctgcagctc agcctctgt gaccactat gtcaccatg
81301 cctgtgctct gctccagggt cctctctggt ctgctgcca ctaaaagcat gagaactctc
81361 ctcccatccc agccctgaaa tatgtcagc ctcaaagct ctggcacat cactgtcc
81421 agatgtcctg tgccctgtga gatattaacc acaactctgc tagccctctg ccatgaggtc

```

81481 agctccgctc ccaggcagg ggaagagctg ccttggggagc aatgatgggc atcgtccctc
81541 tgtctgtgtg ttgggtctct cctccctcc aagtcttggg ctttggggccc agaaatcaact
81601 ttccctccct cctctcttag cacacaaagg gcaggacttg cttctcagggt cttgggggaag
81661 agggagcgcca atgtgtccag ccacccccc aggggtgaa caggagagtg accatcaaat
81721 tcacagaggaa atcccaacct ccaggccctg actgaggcMc taagacccaa gcaagcaata
81781 tcaatgaaag ccagaggtgc ttaaggaaat gaacaaagaa gactgtccag aggcctctga
81841 taaaagatctg tttaaaaaga ctttcttctg tgcggtcatc gctgtgaa tc ccagcccttt
81901 ggaagtctga gggcgtagga tcaactgagg tcaagaattc aagacagacS tgggcaatat
81961 ggttaaaacccc tgtctccata cacataaaaa aaattagcca aacgtaatgg ggtgacactg
82021 cagtcccaag tactcgggag gctgaggtgg gagactgac tgaccttggg aggttgaggc
82081 tgcagttaagc cagatccatg ctgctgcact tcaagctggg tgacaaagca agactgggtt
82141 tcagaaaaaaa gaatttcccc ctctgagtga tgcattggac tgaattttg tacccccaaa
82201 attcatactg tgaaaacctaa tccccagagg gatgatttta ggagtggggg cctttggggg
82261 gtaattagggt catgagggtg gggcctcatg aatgggatta tggcccttat caaagaaagg
82321 ccatggcgag gtgcagtggt ttacacaggt aatctcagca ctttgggggg ttgagagagg
82381 aggatagactt gagcacagaa gtacaaagac agcttgggtt acacagcaag cctcatcac
82441 tacaaaaaact aaaaaataaa aaaaatatgt tgcataggtt ggctcacatc tctagcccca
82501 gctacgtgag aggctgaggc agggaggatct cttgatccca ggagttaaag gacacagtga
82561 gMtaggatcg tgccactgca ctccagccta ggtgatgagt gagactctgt ccttaaaact
82621 ttttttaaat taaaaaatga gaaaataact aatggagact ttttggactt ctttggcccc
82681 tttgcaatgca tgaagacaaa gtgcaagaag gtgccatgtg tgaacacgaa ggttagggctt
82741 cagcagaact aaaaatctgcc agcactttga ttttggactt cccagcctcc agaaacactga
82801 gaaagaaact aagacagtgt aagccacact gtgttggggg tgccttttat ggttagcgcc
82861 tctactgact aagcagtggt actctcaggt agcttgggtt aagccttctg cacttggtac
82921 cagtcgcgtc agtttaagcgg tgcctgccag cttcgggggg ctggggggaga tggagccgat
82981 tgtgtctctc aggttcagca gtttggaaag tttatgagg cttatgagg accgaaataa
83041 aggagacagg gctctgtagg gggctctggg cctctctgtg tcccttgctc ctttctgtgc
83101 aaaaagtga aaggaaaggag gcatcttggg gcatctgtgt cctcttgatt tcccttgatcc
83161 ctgagtcaac ccagacaaa aaacccagg ctctctgtat tgcagggtgc tgcagaaaaa
83221 aagagcaaac ccagacagctg aagagtccca ggcagccatc ggcagccatc gctccagca
83281 aggttagact tggctctctg gaacggcgtt gttaagagac gtaaacatctt caaataaaca
83341 ggttccccct ggcggccccc agggggccct tgaacatctt gtgactctg gaacagcca agctgtctgt
83401 ccatgtcttt ggagagactt tggggaggct gtgactctg ctctgggttc ctctgagctt
83461 tggcttgtaa caacagagca cttggaactg ctctgggttc cctgaaatcc ctctgagctt
83521 tgaaggagag cagacgggaa cttggaactg cctgggttc cctgaaatcc ctctgagctt
83581 aacccagagc Rctctccag ggaactgagt cctgggttc cctgaaatcc ctctgagctt
83641 ttaaccccca gctactgtgt agggggcagg gctctgtgg ggcggcagcg taagcagaag
83701 ggaagtcaac cgactctgct ggaagggcagt gactgtctg ggaaggggct taagcagaag
83761 gccagtgctc tgaatgcc ccaggagcgtt aggtgctg ggaaggggct tctacaagc
83821 ggtctctgca gtgcccaggc aggtggtgct aagcagaagc tctacaagc
83881 cagggtagggt ccaactgaaa ggaacatgc cggaaaggag accccatgtg tgggtgggtg
83941 agctgagggt gctctagaag ggcactccag ctgaaaggag agccatgtg tgggtgggtg
84001 aagaaggcca aggaagccca ggaactgagg aatcaagtga tggagaagga aatggaggag
84061 cagggggttcc caggggcagg cgctgggtcg tgtggggctt ggtagggtgc ccgggtgctg
84121 ggtgcttagt tctgagctga agcaagcagg agccactgtg gatttccag gacaRgacag
84181 acatgactcca gcagaaggca tcaaaagatc cctgaccagc gacggcccca ggaacctcc
84241 accatagtcg aggatctct ttaacagaaa tgcagactta ggcagactta taaggagcag
84301 aagtggatga atggttgctt gaaaggtggt gggggagaaa ggaagcagca taagggtgct
84361 ggggtctttg ggaaggaga agatgttgtt aatataaaa aaaccttga ctaatYact
84421 ttaaatgggtt gacctgtata gtatgtgaa tatatctaaa taagactgtt agagaaaaa
84481 atcccccttg ctgtgttag gactgaggg ggtggggagg caggagaggt tcttgggggt
84541 tccccaaagt atcagaagag gacaaaggct tgacgcaagt gacaaagctg taagtggtga
84601 gaaagtgtca ggtctgggtt gattttgaa agaaaacctt caaactgtt ccaactgttca
84661 tctctcaragt gtgagaggaa gagagggttc aaccttaact ccagagctttt gggctgtcCaa
84721 atgggaaggaa ggtcagcg tgacaaagac tggggaggc cagaggttga gctcgagag
84781 ggtgagcagc agttcaatga ggggcgtgct gagggtgac ttrccaaatt aggttgccaa
84841 gtagagagcc ccttggaaac ttggatttga agagtgcaga gacagatgag tgcagactt
84901 tgcgtgagtc agtagcgttg gcaatacaga gacatggaa ggtggagggg cagtggaa
84961' ggtgactttg ggtgcaggag agagatctgg gctagagagc caaacttggg accattttca
85021 tggatcatata agcaacaaga ggaatgagat cygccacaga ggggtgcaaa gggagacaaa
85081 ggaagcgtcca gacacagacc ctgggcgcat caaccccaga ggttggggga aggaagaaa
85141 tcagacaaagg aaacggggat gacagtgtgag tgtgtgtat agaccaaaga agtgggtgct
85201 cgttaagcca caccagggga gcagtaggga ctgactgcc agctgctacc gacgRgacat
85261 gaggatttggc agggagatgc agcagcatga tgacctcggg gagggtgcct cttggcagagt

```

85321	ggcacacaca	gaggcctggt	ccaagtgggg	ttaagacaca	gtgagaggag	aggaatcgaa
85381	gatggcgtct	ccacctaggg	gggtgctgctg	agagagcaaa	aagctggagg	agtaggtagc
85441	agacgaggga	agccgtcagg	tgtgaatgcg	gataggaggc	agatgttggg	gggttccaga
85501	ggcgggaaga	gtataaaagt	catctggaa	gactgagggg	gactgagggg	agcagctctgg
85561	ctgcttagcgg	caccttgcgt	gtgaacatca	tcgtaaagcaa	gagcagaagc	aggcgtgagc
85621	tttgttcagg	ggtaaccgtg	ccagcatgaa	gtaggttagaa	agtttcaattt	taccaggact
85681	agggcattgt	tcagtagatg	ttacaaagca	agagtgaact	caatgaggga	acctgcagtt
85741	taagaSaagg	aggtcgggta	aggggtggag	ggaaggatca	gtgggtgtag	agacgtgagc
85801	aaggcggcag	catcaaaggc	ctgcaggttt	gctggagcac	aggtcggagc	aaaagggaaac
85861	cgggaataata	gaaggttcac	agaattgcaac	actgcaacgg	atttatcgag	tatagggggct
85921	gctggcaaaa	gaaggtttaa	ggatgaccaa	aggccaattc	cagtggggag	tcctttttgg
85981	aaaccttgac	agctgtgggtc	atggctgacc	tcattgggaaa	ttttgcaaac	tgtactcgtg
86041	gtcaaatgag	ttcccgtagc	gcgtgtcagt	gcgtgtggat	cttgctgttt	ggggggagggc
86101	agaaactcgt	cagaccccat	gccttgccRc	cctcgaccct	gatacgcgat	gccaatagagc
86161	acactgctgc	ggaaaccata	aacgtcacca	gatttcagagc	ttcgagccca	ctctccccaag
86221	ggttttgtct	ggtaaaaggga	agaggaacag	ggttcaactc	ctctggcaga	actctatgga
86281	tgccgcagaa	gacaaactgt	atgagcttgg	cttaacttac	aaagggatttc	cggtYtYtgg
86341	aaagaggagg	aaactcaaat	ctcaggagag	tttaattggca	gagtgtagct	tgactgttgg
86401	tttccctaat	agatgggttt	agctgcattg	gtgaggtaga	ccaagggccct	ctgtagaata
86461	tgctggagaa	caatgattca	agactacaaa	gaaactggaag	gttgagcagat	agatagagat
86521	gaacctctga	gaaacggagg	ccatgccact	gccttgtttc	gggttgaggga	acaaagcctt
86581	gtcaactgtg	tgccagggaac	agagtgcagg	gtgcctggaa	gcctctccgg	aacgtttagct
86641	gtgtccctca	tcgcaggcaa	tgggacctgt	caccagtggc	cagtaacctgc	tgccccagcc
86701	agccttctgt	aggcacacac	tgccagagat	gacctcacgc	tggttgaaaaa	cgaggagacc
86761	gcagatcaac	ccagccgggc	ttccactgtc	agcaaaaacga	ggcgagttcg	ggttgatttt
86821	ctgctctctc	atagctcagg	gccttggggc	gcctgcctt	ccaaacagagc	tcaaaaacaa
86881	tttgagcttc	aaaaagacat	gaaaaatggc	atgtcaaaag	gcctctgtga	atttcaagtt
86941	attaaaactc	aaagagcaca	gccttcttga	agaaagtttc	tttgggcaga	attcaaatgt
87001	ccctctcttt	ttactgtgtc	aatttgaatt	tatttcaact	tttccaggaa	tgaccaattt
87061	tgacggaaata	gtgtagtcaa	aagagataat	gcacatgaag	tttcttttatt	agtctgagtt
87121	atgtcgaaga	gatttacagt	gcattaaaaa	aaactatagg	caaaaattatt	tgtgtaccaa
87181	ttctctacga	ttttttaaag	ggggcaatgt	ggggaccattc	ttgttgattt	ttcaaacaga
87241	tagccctctt	ccgcacagcg	gaagggcctt	gcccacacca	atgtgtctct	gggtggaatgg
87301	taattctgtg	ccatctactc	ccccactggc	tggtgtgggg	gggcaacatg	cccaacagtg
87361	gtactcaatg	ccctcagcac	agtgaattgt	ccaggggggtg	aaacacatgac	tggagatgaga
87421	ctaatcatgg	gtcttccatg	agatctgaac	atttgggagaa	aaaaaatatt	ccctccgggc
87481	tcagtaagct	cacagcgtag	gcctggggct	gcgttatgtga	agagccaatt	tgagaaaggaa
87541	gtctctcgag	aaagagagca	agtcctagga	ggagggagaga	cgctgagcct	tgacagagctc
87601	gtgtggggcca	gggttgctct	ctgtcttctg	cgctgtgctc	agcaaatYtc	ctttttcttgg
87661	cttaatttga	tttggatctc	taattatacc	ttgaattgacc	ctgacatctg	cagttgtcttc
87721	taccttctct	attgctgtgg	ccctttaaana	aaagattagg	aatMatctg	catcgagaatt
87781	ccacttctgg	aatgggtgg	agaggagccc	catggacaca	ttccctctag	aaacagcata
87841	actggtgaaa	agatttgaag	agtcacacatc	aacaaacatt	taaaagtctct	agaaattatc
87901	ctaaagagac	acaacaaact	agaacacatc	atctattcaa	gattcaatca	ccacttctgg
87961	gaagaaaaat	ccagtagaag	acagacagaa	tcgggtggac	ctgagccaat	acttgctctc
88021	tcttccccgc	tcttagctct	tcagttcag	cttatgggga	gttccaactg	gggggtgtgt
88081	ggccaagaac	acagggctct	ctgttctctc	agatccaggt	caaggaaatc	agtaatttca
88141	tcctctcagat	ggggcacagg	ctgttctcat	catctccaac	tccagtgagc	agagggtaac
88201	gtgtgaccga	tggtgaccga	Raggccggag	acacctcttc	cccacagcc	taacctgaggg
88261	caggagctcc	gctcagggtg	cagcaggggc	agagtgctgg	gctctcaatc	acctcaagtc
88321	ctaccttgctt	ctcagaaaaa	tattccatgg	tagggggagg	caaaccaaga	taacagagag
88381	ctacgtccgc	cccaatacct	ggagtgtgtg	cttcagtgaa	gattctggcg	gcaagaaatt
88441	aagaggaggc	tgacagctgt	atgagacaaa	caacctgaac	cacagacagc	cttatttact
88501	tgaaatttcaa	aaaaaaatca	ggagaacaga	tcgttatgat	gatctcttcc	agggggagaa
88561	tgaaatttcaa	agactggcct	caaaatctac	actccacaaa	tttaattcga	tcccaattatg
88621	tgacaattta	tgctctcagg	catttgttga	aataatagag	caatcatccc	tcaattatg
88681	agacttagta	ctgggggtgt	ataccaaatg	aagcagagag	ttatcacagc	agatcagaga
88741	aagggaacgt	catagggagc	cctgtctgaa	tcactgtcat	cccgggagta	ccctgcacac
88801	gctcaagctg	gtgcctctca	acagctgagg	cttccactg	tgaagaaatc	agagcttgtg
88861	aaaaacgtct	aaccaaatca	ggaacaaat	aaatacaaaa	acaactgcaa	acagctctgac
88921	agatggggga	ggggaatcag	tatgcagagc	tgccacatga	cccaaatgt	cagcttctac
88981	acaacagaaa	cacaatagac	tcattcacag	aaaaaaagcg	ggccaacagc	actgtctgta
89041	agaaaagccca	gatgtcaagat	caagactctg	tcagattttg	gacgaacgat	ttcaaacgag
89101	ccatttggaag	tatgtctttg	aaaaaaggaa	accatgctta	tgtaaaaggag	tgtatgatgac

89161 aatatctcat caataaaga ctatcaataa agagaalla taacaagaa ccaaatggaa
89221 attctggagt tgaagaagcac cataactgag aggccttaaca gtagatcgga actggcgagaa
89281 gaagaagatct tgaanaacaga tcaatagaca ttagtccaatt tgaaggacac aggggaaaaaat
89341 gaatgaagaa aagttaacagg aacttaagaa aatgtggaa accataaagg gcatgaaacat
89401 atgcacaagt ggagtccagc aaagaggcaa gtgagtataa agggagcagaa aagatgtgaag
89461 aaacagtggtc aaaaagctcc tcaaatftgg tgaaaacatc aacctcagtg tccgagaaac
89521 tgcagaaact tcaaacggga taaatgccaa gagatccaca cccagagaca tcatagtaaa
89581 aatggcgagaa gccaaagaca aagagaaaat tctgaaagca ccaaaagaaa aatgacttgc
89641 cacatatgtg ggaatccagc tgggatttcc agcgcacttc taataataga agccagaggg
89701 cagcagaatg gcatatccaa tgltaaaaag ctgtcaacaa agaattcttc tccagcagaa
89761 ctatttttca aaatlaagtl gcaatataga cagtcccgaa taagtataaa cagagaattta
89821 tggagtgaacc cattttatca gacatactaa aggaagtctc tcaagctgaa agcaagtga
89881 ctgagttggt aacttgaatc cacaccacaa cacatacagt accagtaagag gtaattcacg
89941 aactttaaaa gacaatataa atgggtattt ctctctttt ctcccttaa ctggcttaac
90011 aagcaaatat atagaagaag atglataata ttgtattatt gggcccatat aatatataaa
90061 tgtatacat ttgactaaca gcacaagca ggtgaacaga ggtgaacaga ctggagttaa
90121 aaaaagcac cagatgggaa ctgtaatcca cgcacaacaa tgaaggagac cagaaaagt
90181 caacaggagc gcaaacgtgt atatatata tactctctc tctttttatg tacttttatc
90241 gcatgaatt tgcataagca ataatataca caatttctat caatttctat gtttgtaga tatatagctg
90301 tgggtggagc agaaacttgg agtgcctcact cctccaaag atgtccatcat aggaacctac
90361 tatctgtaa gcttatgtca aaggaaaaat aacattgtaa atcagatgac tttaaaat
90421 tgaattatc ctggattagc caactggggc cagtgtaact caaagggttc atakaggggg
90481 aaggsgaagc caagaagag aaactggagg gatggtgca cgggaaggag tgggctgat
90541 gttgtggcca taagaagag aggcagagag tcatgagcca aggaatcgag ttggctccga
90601 gaactggaa agggcagagc aacagatctc ccccgagga ggaacagaa ggaacaaaac
90661 cctgcacaaa cctgtattta aactcagtc gactcatttc agactctgac tcccaaat
90721 gcaaaaat aacatgtcat tctttaagcc actaaatttg tggtaatttg taccgcgagc
90781 aatcaaaaaa caacacagat tgcacagcaa gaacagtaaa caggctggggg
90841 ggggtggctc acgctgtaa tccagacat ttgggaagcc gaggcagggc gatcaaggg
90901 tccggagatc gagaccatcc tggctaaccac ggtgaaccc gctcttact aaaaacacaa
90961 aaaaattgccc gggcgtgtga ggggcgcctc gtatgccag ctactcggga ggctgaggc
91021 ggggaatggc tgaacccggc gagggggagc tgcagtgag ccgagatcgc gtcacttgca
91081 tccagctgtg gcgacagagc gagactccgt ctcaaaaaaa aaaaaaaaaa
91141 aaaaaaaMam agatgggggK ggaagaatag agctataaaa gagttaattg tctatagctc
91201 acaggaatta agttaataa aatctaaaga tctgataag atgcatacga aggggagag
91261 gacatcacca aagtacaga atagaggact ccaaaattct gccctccac aaaaaagga
91321 aataagctag taaatctcag aatcgactaa tttggaactc cagaatttaa cagaaaact
91381 acaacacacca gggaaatgctt aatgaagaaa gcagctactg aattttgggt acagtgggt
91441 atgggtgttt aatttacttg ctaccatta tcttattct caactcagca gttggcatga
91501 agatggcaga ccaatctctc ggtgtaggtt gctgatacca acctacacat ccaaggagct
91561 caatgactcc cagtttgaaa actcaaaaaga aagaatatg gacctatat cagaagaatt
91621 gtgggtgtgt gttttaacc cctctctggc ttgctttgt tccaactgtc tggaggtctc
91681 cccagagctg aagtggtrtc coagacagtg ttgtcaaaa grttaaagg ttRaagtga
91741 tagctgcaga caccctggggc acaggaacat kTgggggca agacRcaaa taacatcat
91801 taaaggcttg gggaggaaga ggttagggaa ggagatatt gaggaataaa gtttttggga
91861 aagctcttgt gcataccaga gaactagga tcttaccat gacatgctca gggtagagtg
91921 catgctcaga aaaggtctaa gaagactcaa aagttrtca ccttggtcta attttaggt
91981 ctgtacagat aggaagtga ggctaagaca gagtgtaaa cagctcggtc aagcatcaaa
92041 ggtgtgctca aacacagaa cactctacaa agacgagag agtatttctt ctactcctg
92101 cttttaagaa agtctctgtc aaatcactag ctgaccacta agctaagaaa acagagactt
92161 tgggtggcat atataagaa caatgcagac tracaacaaa tagtttgaaa aactactaa
92221 acaaaacagc aacaacacat aaaaaaacac aagccccggg aagtgggggg cctctgatt
92281 ccaagagtct ccaattcaaa atgtccagtt tccaacaaa atrcacaaag atgcocaaat
92341 acaaaagagg atgttccact cacaggaaat aaaaaagacc tccctgagg aagccttagac
92401 acctcatgtc ccagacaag attttaata aattgtctta aatgtrtca agactctaaa
92461 agaaacccat ggcaataaac gtaagttaa cagaacacata atgtatcaca aatatagaaa
92521 ratcaataaa gaaatagaaa rtataaaaag gaaccacaaa caatttrgga cgaagaaagc
92581 etaataactg aaaaataaat tcaatagat cagtttcaata gcagatttga gcaggcaaac
92641 aaagaatcag caaacttaca tttacagcaa tcatcccaa tgaagattga ggggaaaag
92701 aatgagggaaa aaggaaacaga acctaaagaga tcaagtacac caaaatacaa aaaaatgaga
92761 gttcaagaaa gagaatagag aaacaaagg gcataaagaa tgrttgaaa acaatggctg
92821 ccaaatctcc caaatttgat aaaaagaaca tgattctaca catccaagga gctcaatgac
92881 tccaagtttg aaaaactcaa gagttaact aagataact gtaactaac gtcttaagag
92941 caaagacaaa aagagagctc tgaatccac gagagaggag cagcttgcca catatacag

93001 ggctctcaat gagcttaaca actgatttgt catcagaaac aataaaggcc agaaaccaac
93061 gggatgacat atttaagggt ctgaaaaaaa tactgtctctg cggctcaaaaa ttccaacaaa
93121 acctctcttc aaaaaatacW gaaaaaatWg acactctcag agaaacaaaa actggagagag
93181 ttgttttttc agttgacctg cctcccaata aggtgttaag aggtctccct agggtgaaat
93241 aaaaggcagc aagacagtaa ctcaaggcca tatgatgaaa tgaagaatac taataaaggg
93301 aaaccccaaa tcttcgtttg taactcttct ctltttttct atgtatgaaa gtctatagat
93361 gggtaagca ataattataa atctacatta atagacacac atgtatgaaa tgaacaata
93421 ttataataaa ggggtgggaa gacagagctg tataggagca agccattttg tatactattg
93481 acactaagtt gatatacaat caaaccaaat tggttatgtgt ttaagataat agctgttaat
93541 cccagagtaa ccaataagac catagctttt aaatatatag caaaataaat gaacactatg
93601 atcaaaacag tacactaaaa aacaatgtgt taacacaaaa agaaggaat atttgaggaa
93661 ttgagggaata aagaatatga catatacaaa acaaatagca gatagcagaa ttcagtcctt
93721 ccttatcagc aattatttaa atacagatgg actaaaatgct gcaattaaaa ttccagagaa
93781 gggagaatgg attaaaaaat cccaagatat aatccaacta tatgcttctc cccaggggac
93841 caaattttaga ttcaaaatcc acaaatagtt tgaaaaatgaa aagatggata aagaanaatgt
93901 atgcaaatag aaaccagaag agacctggga aatatcagac aagatagact agatagact
93961 tcaagtccaa aaatttttaca acaagaaca gcattatata ttgatgaag agtcaatcca
94021 ccaagaagat ataacaatta taacataaaa caacactaac aacagaatcc ttccaataa
94081 gaagcaaaaa ctgacagaaat tgaagatgga aatagaccat ttccaataa ttgctgaaga
94141 cttcaaatatt ccactttcaa aaatggataa acaacccggc acatgtcag taagaaaaat
94201 gaggacatgga acacactctg aaatacaact gaectgcagc acatacagat agacccccac
94261 tcaacacgga cagaatgtgt attcttctca tgtatgatga tcatctccca ggaatggcca
94321 tatgtttaggt cacaaaaaaa gtctcaataa ccttaaaagag actgaacaca tacaagaatc
94381 cttctccaac cacagtggaa taataactaga aatcaatgac agaaagaccc ttttaaaat
94441 tccaagaata gtgaaatatt aacaataatc tcttaaaagaa ccatttaggtc aggttaatat
94501 ttacaagatc aattagaata tataataatct ttaaaatctt gagacaaagt aaaaataaaa
94561 cacacacatc ccaaaactca aaggacatac aaaaagcagtg cccagagaaa tattttatgc
94621 tccaatt tcc tacattaaaa ataataaaga tctcaaatca ataacctaac ttacacatg
94681 aaggaacacag aagaaaagca acagaaactc aaagctagca gaataaagca aatatcaag
94741 attatagtg aggtagacaa atagagaat atcaacaaaa ctgacaacaa taagaagact caatgaac
94801 aaaaagctggg tctttgaaag atcaacaaaa ctgacaacaa tttagccaaa ctgacacga
94861 taataaagaga gaaggtctca attactaaaa aagtgaaaac aagtgaaaac ctgacacga
94921 acctttataga aataaaaaatg attataaagag aataattatg acaatgtgat acgaacaaat
94981 taataacact agaagaaatg gacaaat tcc tagcaacaca gaactacca aattgtactc
95041 aagggagaat ggaataattg tacagatcta taacagtaa agagcttgaa tcaagtcatca
95101 aaaaactcca ataaagaaaa acctaggacc agatggcttt actgggaatt ctactcatca
95161 cttaagaaat taatatccat ctctctatga ggccagcatc acctgtgac cgaaacccga
95221 caaaaactc ataagaaac tagacactaa tgcctcttat aaatataggt gaataaatcc
95281 tcaatgatat tctgcaaac caaattcagc agcatattaa tcaactatca tacaatccac
95341 aagtggaatt tatcgttga atgcaagagt gttccaacac agcatataaa atccatataa
95401 tacactacat tcataaaatg gagaaaaatg aaccacacag tcaatttggt tgaatcgag
95461 aaagacttag acagaaccca acaccccttg tatgcaaaaa atactcaaga actcagact
95521 agaagggaaa tctctcaga tgataaagg ttattacaga aagccaaaag tactcatcat
95581 ggtaacactg aagactggaa gctctgcttc accactgtca ttcaacatg taccggacat
95641 tctagttaga gcaattcagc aagaaaaaaa taatatataa aaggacatcc aatttgcaaa
95701 ggaagaagta acactctctg agggaaaaatg gggaatgatg gctctgtgggt acagggtttc
95761 ttgtggagc aatgaaaaatg atctaaagtt gactgtggag actgtgtcac agctgtggg
95821 atatagtaaa aattatggaa ttatgcgcc gaaacgctga actacagcat atgtgaaat
95881 cagctcaata aaaccccat aaggaagaa atactatct ttaactctgt ttactctgtt
95941 ctacagacca ttccagccc agcccaggcc ccaggcaacc ctctgtgggt gacgcagcat
96001 taectgtctc ttgctgtacc agtgttgtgt agcgaaggc ctgtggaggc agctgtgtg
96061 gtacatcggt aaccaagaaR gggagcaagt agttgggtga caaacctgtc acatctgtc
96121 cttaactttc aaacctggct gatgggaggc caatgtSacc ctctaagcc tcccatagc
96181 agcctcgctc tgtcttctga cagtggcact gtccacccc aggcatacag catctctgcc
96241 gcaatttgtc tccgcagtg gagccccag gagcaacta gacaggagag tgaactcaga agtccagag
96301 atctctctca ctMtagggaa ggaagaaagc ctcttcagct tgtctgtcct cggcaaggc
96361 aaagctctca cgaatttgtc actcctccc accaggctcc agccggcacc tgaactctag
96421 ttgacagaaa gagcccaatt ctgcatcagg cgggctgggg ttcaaatccc agatctctcc
96481 ctatctcaat ttgcccattc ttgagccttcc atcccttacc tagaaaacga gttctttcca
96541 cccatcgacct tgtgagacag cgtaaagtga aatgacatc acacggtaet ggaagcagag
96601 ctgctagaaa cccagccagg cggggtttga acttgttatc tccaaaagct ctcaggaatc
96661 gtattttctgc ctgagtgtga cctacttcag acaggaccca atgtagaat tactctgcca
96721 ccaaaaaagc aggcctcaga cccctccat cctcttgaat ttaaaacaga caccacacg
96781 ggccagtgc tgaggggtct gcagatgaga gtcagaaacc agcagccccc caccacacg

96841	agacagggtcc	ctgctaacta	gacacaagaa	ttggcccttct	gggggtgtctc	tgcgtgttct
96901	gtgtataggtg	acctagccccc	ggccctccag	gagggacact	acgtgcaggga	caattttggaa
96961	acacagtcctg	gtgttctctc	aaaaagtgaa	acactgtgtt	atcatgtgag	cgacgaattc
97021	tcctatKgat	ataccataaga	gtaatgaaca	tatatgtaca	cacaaaagact	tgtccacaaa
97081	tgcccatcga	ggagaagctg	gccacagaga	agatggccca	gagaggaggt	gctgtgtgca
97141	gaagcagctc	tgaggagaaa	gacgtgcaca	ggccaaaatc	agattactga	caactttttt
97201	ttaatgcaga	gatttctgtc	agaattacct	atttttgact	tactcttgaaa	caactgaggt
97261	ctgtcacaac	tgaccccgaa	tcttcttaac	agccaaaagt	atcagaactg	agaggttttS
97321	tgtctcactg	attttttttt	cttttctttt	tcttttcttt	tttttttllg	gatagggtct
97381	gactttgtca	cccaggctcg	agtgcaigtg	tgccactctc	gctcactgca	accttcaact
97441	cccaggctca	agcaattctc	ctgcccacgc	ctcccaagta	gctggttcca	caggtgcatg
97501	cccacacgtg	lagctaattc	ttagattttt	ttggtagaga	tggggttttca	ccatttttgc
97561	cgggtctatg	tcgaacccct	gaagctcagg	gatccactcg	ccctcagactc	ccaaaatgct
97621	gggatatttg	gcacttttgt	gcacctggcc	tgtctctact	gatttctgat	cttallttgt
97681	attacttttt	tcttcttatt	tctcttgaa	caatgttgc	ttttctaac	ttctcaact
97741	ggaaatttat	ctctttaatt	tttagacctt	tctcttttct	aagtgtaaat	tccactcagc
97801	tttgtatatt	tagtgtcttt	attatcattg	agttgtattt	ttgaagtctc	actatatttt
97861	ctctcttgac	ccatgaatta	tttaaaagta	tttttttata	atttccaaat	aattggggatt
97921	tttgtcttgg	gggctattaa	tttctctctc	acaggcatga	tggtcaggga	acaaagacct
97981	ctgcacattat	tcttttagtt	tgtgtgaagt	tgtcacttaa	ttgaccagtt	gttcaaaagt
98041	tttctatggt	gcttgagaaa	aagaaataac	ctctacttgt	cagatgcmaa	attctcaaat
98101	gtcaagactta	tatttttgtt	tatttagaac	tctcttatct	gtgctggggt	ttcttctaac
98161	ttgatccaa	accaattgga	aaaaaatgtg	tggtattctct	caaaaatgat	gataaatttg
98221	tcaattttctc	tatttttaac	aaagttcacg	tcaaaatagt	ttgagttatt	tttlaattgt
98281	actctacactt	tagaactgct	ttatcatctt	ggggaataga	accttttctc	atttaaacgt
98341	gacctgtctt	tcccacgacg	cgctctctgt	cttgagttgt	tgtagctgtc	agattgatac
98401	agcaatgtgc	gcttgccctt	ggcgccctga	taggtctgtt	tccatacttt	actttccact
98461	ttttcttggc	tttacaactt	actctgttga	tagcgtccag	ccaaattttg	aatttggggt
98521	ttgtctctgt	tgtcagatat	gacagcctct	gtctttgaa	ttgttgattt	aatccatttt
98581	caatttaactg	gattattgat	aaacttgggc	tgttttctcc	ctctctactc	tatgatctca
98641	attttgtctg	gtttttctgt	ctttttttct	cttttataat	ttgtcttata	aaaaaagca
98701	tcccgccggg	gcScagtggc	tcaggccctat	aatcccagca	ctttgggagga	ccaaggtgtg
98761	cggaatccct	gaggccaggga	gttcgagacc	agcctgacca	acaggggtac	accctccctc
98821	tacaaaaaat	acaaaaaaat	tcgggggggtg	tggtggacca	cgctgtgat	ccagctgtgt
98881	agggaggctg	agggcaggaga	atcactggaa	cccgaagggc	agaagctgca	gtgaagccag
98941	atcgagccac	tgcaactccag	cctgggtaac	aaagtggagc	tccactctca	aaaaaaaagc
99001	ggcatccctt	tttctttaat	ttgtctctgt	accactgtgc	tggtctcccc	agccagggtg
99061	caaccctggg	gacacacatg	cccctccagc	agtgaattca	gacctctctc	acccttttgc
99121	ctcaggctcc	tcacttctgt	cccctctctc	tgagagggtga	caagcctatg	ccacaccaac
99181	tccaaattctc	atcatcggtc	ccagtcacgt	ccctgcacca	gctgctcatg	ggcctttctc
99241	tctgggggtg	ggctgggagc	tgtggttctc	gctgctgtgt	ccctgcaactc	tccacattgt
99301	ccagttctga	ggtagtcgga	gccccctcta	atatacgca	tcccaggggc	ggcgacagg
99361	gcacaaggcc	acttcagttg	caacacggat	ctcactgcaa	gaactaggag	gggctccctc
99421	taggaagcag	caactctggac	tgtttttaacc	acagtcccca	cttctctctc	ctgtcttgat
99481	gtcccagact	tcatacacac	acgtctgtgc	ctattctgga	attatcgagt	attattattt
99541	tatatattctg	tcgctgggca	tatgatgtct	ccatgacaag	gggcccgtga	gcttttagat
99601	caatctctag	aaggactcca	gccagacccc	aaaggccaca	tattatcaat	tggtgtttca
99661	tgaagtatac	agaataggca	aatccaaaag	caacagggag	agattaKtgg	tcactatgag
99721	ctgggaaagg	gaatggggag	tgactgtctc	gggtgcaggg	ttcccactgt	gggtgatgaa
99781	acatttttga	actagatggt	gatgatggct	gcacaaactc	gccaaaaaac	acttgattac
99841	acactttcat	actcatatgt	attatagttc	attgattttt	aaaaggtcac	aggaagcaac
99901	tcaaatgtct	gtaggaggtc	agaagagccc	cagatgggtg	aagtcggggg	gcttttagat
99961	ttaagcccaa	aagaatatct	tctaagctac	gactgcacac	acacactcYc	tctcactgtS
100021	ctgaagccc	atagccccc	ctgctgactc	ctgtKgatgg	ctgtKgatgg	cgctcaaaagt
100081	gcacaSaaat	atattttggc	tgaaagaaac	acaacagggc	acaaatgagg	agggctgccca
100141	agcagcctcat	tagcagcatt	tgtcatggtt	ctctggggag	tctgcttttt	cccccgacSa
100201	gactcttcag	ggacagagac	cagattgttc	atactcagtg	cccagcaggg	cgctggcctg
100261	gggtgtctgc	agacactgac	tgagtgaaat	tggatggatg	ttgggtggat	gaatgagttg
100321	tagatgaata	ggatgatgaa	tgtatggatg	atggatggat	gatggatgaa	tgaatggata
100381	ataaatggga	tgatgggatg	ggtagatggg	ggatgggatg	atggatggat	aaglagatga
100441	atggagatca	atgatgggat	gatagatgga	tgatggatg	agtggtgtag	cgatgaaat
100501	cacagtggat	gaatgaatgg	ataaatgatt	ggatagatga	atggatggac	ggatgacaga
100561	tgataaaatg	gaatgaatgat	tgatagatg	gatgaatggg	tagacagaaa	gactctgata
100621	ctggaaagtg	aatccttcac	atccatggca	agtgaaccgg	gtaacaaagt	cagaggccaag

100681	accatcagcag	cgtcatttgt	cactgtttgcc	attttatcac	attttttggg	cagaatcggt
100741	catgcttgac	cacaggtcat	ctctagagg	ttaggattgg	gggtcaggaa	ctgtgggggt
100801	ggctcagggc	taggctgtcc	gatgtggct	tgtgctgtg	gagcagactg	ctgttccatc
100861	agctcatccc	ccatctctcc	ctctgtgctc	cacagtgtag	cttcagatct	tctctcaatt
100921	caatcttggt	gtccctctac	ttttctgtcc	atctcccaag	acctggtctg	ggctcagctc
100981	agcatggggg	cagcaaaagc	ctgaggtaga	ggctctctcc	aaagtattgg	cttggtctgg
101041	ggccgcctca	cctccattca	aaggacacca	tgtgaaaaac	caatgtgccc	ccaaagctgg
101101	gtgcgcacgc	acagacactg	atggacaaag	ggccattctc	tgtctctctc	ctgggtctgc
101161	atctccaggg	atcaggggct	ctgcagcagc	aggcagggtg	gggcagcctg	gaagtgtgtg
101221	acacaaattg	ggcaactgtg	ccctactgga	cctctttgca	agactgtctg	gggcmftaga
101281	aactgggtgc	tcaccaactg	ctgtctctgt	aggcctgtgg	gtctgctctc	acccccaaac
101341	catgagactg	ggcaggtaac	caccagttct	YttctStgtc	ccacagctct	cccgagctt
101401	ggccttggtc	caccgtggac	agcgtctctg	acacggctYt	tcttcccaca	gggagaccgg
101461	gagggtcttg	ttctgcccct	agagcagaag	ccccaggctc	cttcagggtg	gagcacaggt
101521	gcctttcgca	cagggccacg	cagccagctc	agccccacga	acagttaagc	ggcaggctcc
101581	aaagtgttcat	cttgtgaccc	ggcctctagg	acacgggtgc	ccagagcga	agaggagctg
101641	ggcttattat	aaggaaaata	gtgtcaggcc	tgcattggaq	ctttgggctc	gaggcctcag
101701	ctgagagcaa	ggcagcagga	ctcatccaga	gacaggacag	acctgggctc	ctgtgtgccc
101761	agctcagggg	agctctctgg	tgcggctcgg	tctactccag	acttctctac	atggatgaar
101821	atgtgtgatt	ctgtgctctg	acctatgggt	tgtgcatgqc	tcgtgtattg	tgagagagag
101881	agaggagaga	gagacggaga	gaattttatg	acgtgttaact	acccaacaga	gtccacagct
101941	ctcacactcc	actgcaaaac	aggcaggcgg	ggccagcatg	agtggggggg	ccaggcagga
102001	atcggcaggg	agggaggggg	tgtgttcgcc	catccagagg	aggcagcag	ctctccaccc
102061	atgtgccttg	cgaggaacac	ggccagggtg	ggcaggccctt	ctgaaggaga	tgaanaattg
102121	gattttctat	tgaacacctc	tgaattttta	gtgtacacct	aacatttctg	aagccgttgt
102181	acgagctcaa	taaaacattg	cgggggtggg	ggccaggcgc	ggccagagct	cggaatttgg
102241	agggctggct	ttagtcacag	acccccagt	acatacagag	gagagagact	aggctcagcc
102301	agagcacaga	ctccgcaggg	caggccccca	ctgctggccc	cggtcgctct	ggccatcact
102361	ggggctacca	cagccccatg	ggagcaccaa	gSaccacca	aaaggagggc	ccaggcctgg
102421	tcccgggggt	ctctgtcacc	aaggggccac	atttgtagtt	tcaaagggtg	ccgagtttgt
102481	tctcacagcc	ctgtgagtcc	ccacactcct	tcccagttca	aggcctcggg	acgaatacca
102541	ggccagcttt	ctgtctgtgg	tgtatgaagg	ggagactctc	ggagactctc	gggatattgc
102601	cYAKaaca	ggagcacagg	cactgagct	tctggacaca	agggcagaa	ctctctgaga
102661	ccatctcgg	ccgtttgtct	tcaggctctg	ggggcctctg	ggagactcca	ggccagggtg
102721	taggtgtccg	gagcccgggg	aaacagaact	cagaccggga	actgggcacc	agatgcacca
102781	ggccgggggt	tgtccgcccc	ggccggggag	tcagggggct	gctccgaacc	ggcgccacg
102841	ctgtctccct	cagtgcgcga	tcacaagct	ctgtctgctt	gggctggggc	ctctgRccct
102901	cccggctgct	ggctcccgc	gatgcgtctc	agatgggctt	catcgctctc	tccaaagaac
102961	caggagcgac	ggcaggtttc	ggctcacaga	aaagggaacc	ctggaaaggc	ctctcttgac
103021	atttccgaaa	ggaaaaataa	aattcctctg	aagtacattt	caactccgag	gcagccctgg
103081	ccacatccaa	gaagctcggt	cctgagcgcc	gtgggctcca	gtgaaggctg	ccagcactga
103141	ggcgccctcc	tgaccoccat	cccgcccact	ggcccgggct	tgacacacct	tacacggctt
103201	tcacaaattt	tgaatttagct	ggcaacagtt	taaaattga	agatttgaag	taaaatacat
103261	ttccagcttc	tcttataaaa	tggaggcccc	tctgtctgt	ggccagcccc	gtgtctactc
103321	gggcccaggt	atctctgctg	ggctcactgc	ccaggccccc	atcagttgtc	ccatcagaga
103381	ccactgttct	cagcgaaagc	ctgaccccc	ctgaccccca	aaagttgaga	ggacagatgc
103441	tctctatctg	ggggctctgg	ccttctctgg	ggcagtgtaa	tcacagctgt	ccctccagag
103501	ccctctgcac	acgctggggc	tgggactgct	catgctgtgt	ggctctgatt	tggtgtggga
103561	gacctgagca	ggccatcatg	cttcttattc	ccagcctgac	acccccagac	cgccccccgc
103621	acacccagac	aatgtgtact	gaactcctgc	tcaacttgac	ccctgctgac	ctccccgcg
103681	cccaccaggt	gctcctgctg	tggtttgggc	agcacctagt	ccttttgcgc	ctgtggagcg
103741	tgcgctggcc	agacttgacg	agccatctca	ggcctttgtg	ctctgactc	ctccctcca
103801	agcctggcca	ccccagcctc	ggccccagct	ccctgccaag	gaattgtcaa	aggcagagaa
103861	taggagagct	gaaggcagcc	cccatggcac	acccccagga	gtccctgaga	gtgtccaaaa
103921	aagctctaa	cctggggcag	attgctcgca	gctaggaatc	tgacagctga	ctcaaatatg
103981	tgtccagaat	gatctgggaa	aggctcctga	gaaaaggccc	agacacaggg	accttccag
104041	ggccggatcc	cacagaggaa	atgattctag	cgagcccgca	ctcgacaggt	ctlgagttcca
104101	gacaggggtg	gagctgtcac	aggcagccgg	gtgggctctg	gatctgggtc	cttccagctg
104161	cacgtgtatt	tggaaatgga	atacggttct	ggcaggcaag	gagtatcccc	ctctgaccaa
104221	ctcttgagct	ggctaccagg	actcccaacc	cccttctctc	agatggcgcc	tgaggggctg
104281	atccggcgcc	tgctctcatt	cccatggcca	cctgatccag	ggcgagatct	ctccctgcac
104341	acaggacagc	ggctcctgct	cggtccctgc	ccctctccgt	gctcaccttc	ctcggaaccc
104401	agaccatgcc	acctgtgcga	tggcacagcc	ctctctccac	gacgagcccc	acaggaagcca
104461	tctgtatctc	acctccagct	ccctcaggcc	ctcaccctgc	ccctcctgct	acacccccaa

```

104521 acaagcccgc tcctcacccc tcaggctact cgcattctgct gccctggccc ctttctctcc
104581 tggggaaactg caacccacac tcaggctctc agcccaacgt ccagtgaactg cgggctcact
104641 ctgtggcact tcocaaagtct aggtatttgt ggactgacta gtgcactgtgc ggactcactg
104701 ggctgtatgat ggtcagggtct gtgcctgtctc ggctccagag cctggcccag tgcccgggtcc
104761 tgcgtgtgat gcggggtaaa gatctgttgg atgaatcaag ggagagaggaa agggggacag
104821 cccctgggatc ccacaggctgc ttatgttcca gagacaacgc ccaggggaaac aggggtccaa
104881 gcacgctggg agcagtgctc cagcccgagc gcatcccag ccaggggcag ggatctccg
104941 ctctgtctgc ctacggcacc aaacgggctc gtggcagta ccaggggcag ttctctgatac
105001 cgggtgtctc gtccagcat atctctctgc ccatctgcag gtgctgtgc agtgccctgc
105061 gacattccca aataaaggac agacagttaa ctgtcagaaa gtcactgtgc ttgggctctg
105121 aaggaaaggga gcccggtctc actgggcacc cccaaatttg gccactgtgc agtccccctg
105181 gtctcacaac tcggggccctg gaccacagct ccattgaacc acaggggagg agcagagtca
105241 caagaagaag agtgatgaac agcactggcc acttgatttg cagggcacag tgcaaaaatg
105301 aaacgcacag tccttgtctc aaaaagtgat tacagattct aaggcaggaa tgacagagt
105361 ttaaaagtgg ctaaggcccc tgccggatgg gtcaacaagg ggccaaagtc gtctctctca
105421 agtgtaggaa agtgggacta gccaaagccc agggccaccc aatgtcccat gggcccatct
105481 gaccagtagc caccacagc tgcacaactg agtctgcagc cctgacactg tggcaaacag
105541 aggaggctgc agtgcgaagc cctcaaatgc atgtcaggag caactgtgtc cccggagtg
105601 agctcctggc cgacggggct tcctcaggac tatgttatc ctctctcttc agcgatatct
105661 acgcacaacg cgtgtgccag gattgtcca aatgtctga aatctcactg tccgcaaaag
105721 attctccccg agcttaacgt cctaagtac ctggagacac atctccactc acctctcgt
105781 gcccacgtt tcgtcagccg ttaagtggag tggctctgg ttgctctgg aagcccttc
105841 gtccacaaca cagatggggt cacaggctac cgggttttgc cagcccaagg tctgagcgg
105901 gacaggaggg cactgcact ggtgtcagca gccctgggtt ctaagcctgc tctccctcc
105961 tgatgcagga gtcaaaaggc ctggtctcca ctgtttctac aggttactat ctgtgcaacc
106021 tgcggcacca cgttttctc atctgtaacg tgggcacagc aagaacacca acccaactgt
106081 gggtctcata caagaatcct agggatgat gtacctgac cgtgtgtcca cggcggtca
106141 accgtatagt ctccaggaac cgtctacgtt ctcttgtctc ctctgtctga cgcctactgc
106201 tggtagactg gtacccctgc acctagtgtt tctcttgtct tccaggagta cgtgtctat
106261 taaatgacac catgaagagt gccacacttc gcgcaggccc tcggtgaatg ccgtctctct
106321 tgaacgcata cttggtcca acggggcagg agaagatcca ttcagactca ctcacagaaa
106381 aacagacttt tgtgcaacta acgctaaca aactggaaa gaaaggagaa gtacagcttc
106441 tggaaaagct tgtatcgaa atgacagaaa gtcaatatct aaaaaggagt aagcagctta
106501 ccccaatata taagacccca ggagataaac aggcaaaagg cttaggaaata gaagacatgc
106561 agagtataca aaccagggga agatcactca tctctgccag taaccaaaaga accgcaaat
106621 aggccaaaaa gcaacgcgcg tttgctggcg gagcagcaaa atattaacaa ctgagcccc
106681 cgtgctggtg aggtatttga gaaaccagga gctgctgtgg gagccagctc ggcctgtccc
106741 cgagtggtag gtctgttaa cgtctgggat tctcaattca tttagictac cactaggtgc
106801 cccctgatag ctgctcaggg ccagggtctgc agcgcgtctg ggcctggtca gggccgcatc
106861 tgcgggtgac gcaacagata cctatttgca ggaacagaac ctgaggtctc cctccaggat
106921 tctcccaccc ctgagtttca agagagtttg caccaggagc agggagcatt ccaagtgtct
106981 cgggcagctt ttgtgtgggg aacagctggg agcgtgcct cgggcccctt cgggtggaca
107041 ggaaggctct gcctgcaaa gccttagaggc agccacacac cctgccctgc accactctaa
107101 tctagctgag gccacagttc ctgcatccat tcatcccgcg cagtctgggg tccctccca
107161 cagtggcagg caaggggagg ctggaccggg cctgctagac aaacaaggat gacacacagt
107221 ccatgggacc ccaaacctca cactgtccct gacagcccg gcacagagc acagcctctg
107281 cctcccaggc caaaggaggg tggcagtgca tcccagggcc tggctgctg tggggctgca
107341 ttgggggtct cgtccccgc cccgtgagtc tgggaacaca tgggtcttgg ccttggcacc
107401 tcttttctct ggaaaaacac tccctctggg ccagggcagt gcaagagcat ggaagcggc
107461 agctgtggag tcagagtgt ggttgccaat tctgactct gacgtggcc cagaccactc
107521 aacctgccac ccccaccag acacctgcc ccgcccacc accagcaac ctgcggccgc
107581 agcctcagc caccagccag cctcccagcc acgtgaectc tcaaggggcag gatgtgtct
107641 tgatttctct cttcacacca cgtcccggaa gctccacctc aaactctctc tgaacccact
107701 gcctcacccg ccatctgtca ccgggaactc tccctgtctc cagccgtcc cgcgaaaacc
107761 cctcaaaagg cttccaccga cagggaatga attctcacc cgtaccgtga cctctctctc
107821 aaagcgatga gtgccacca tccctccggc cctgcacgc cteccactc cccattagc
107881 cctcagttcc aggtactagc tcttctctct gccctctgtg ccagctgcgg taactccagc
107941 ctcagggtcc ctgcatgtgc tgttctctct gccgtgaaac cctctccag cgtggggtct
108001 tcccatctac accaaactcg ggcgcctccc tcccttacct ccctctcac gtctgctct
108061 caccocgttt tactttgttc acaggattct ccaagtctgt cgcactatc ctagcaggt
108121 tctccctgcg gggctgtgag cccagagaa agggactcga tgggttctc cgtcgtact
108181 agccccagca cccagtctct taagacacat ggtggacaca taqcaagagt ctgacgagt
108241 aatcagagaa aatacacttg tggcagcagg caaggtgcc tctctgagcc tcaagtctcc
108301 ctcaagcagt agagtgcata atgtattat tcccctatga ttgggggagc agtattattc

```

108361	ccccatgatt	tgggggagcag	tattattccc	ccatgatttg	ggggagcagca	ttattccccc
108421	atgatattggg	gagcagcatt	attccccat	gatttgggga	gcagcatat	tccatgatatt
108481	ggggagcagc	attattcccc	catgatttgg	ggagcagcat	tattccccca	tgatattgggg
108541	agcagcattt	ttcccccatg	atttggggag	cagcatattt	cccccatgat	tgggggagca
108601	gcattattcc	cccatgat	ggggagcagc	attattcccc	catgatttgg	ggagcagcatt
108661	tattccccca	tgatttgggg	agcagcattt	ttcccccatg	atttggggag	cagcatattt
108721	ccatgatattg	ggggagcagca	ttattcccc	atgatttggg	ggagcagcat	attatgtagc
108781	attaatgaac	ccgcatggca	ttcaggtgcc	ggccttttcc	ccccaccagc	cagcgccaca
108841	gcacaccttc	cacacaggcc	cagtgccccc	gtcacctctc	ccctgtctct	caggcgtctc
108901	tctgggtgtg	ccgacagttt	gggaaaggag	gcgtgggag	ccgtgagctg	gcctctggcc
108961	gggttctcta	ccagctgtgc	tctgtggcga	gccacacatc	cacggccaagc	ccccagccaaa
109021	gggaccaa	caagagctct	gagagacatc	ccagccgaag	aatccagcccc	ttgtcccaaa
109081	cgctgtgatg	caagggcaca	cgaagaaggg	tgagacagaa	ggagagcaaa	attgaagccag
109141	agagcagag	agagagaaga	cagagccaca	cagccccatc	tgactgcgcg	acctttcccc
109201	cgctcagagt	ccctcagggc	ccagctgcta	gcgtgctcta	ggatataatt	caagagaaaa
109261	gggtgcagag	aaaattggaa	atttttatag	aaaatttttc	agccttggag	cggtctgaag
109321	tctgtccgaa	cctgcagaga	attctgagag	gagatggggc	ctgtgtgcag	agggctccgg
109381	ctggcacaaa	ctggcagaagt	ttggacctcg	gacccaaatg	cagtttctga	gccaggcctg
109441	tcctgcctgc	tgtctgggaa	gagcttctcg	tccctctcag	atttccccat	agctgctctc
109501	gcagccctga	tgtcgttacc	caacgcctct	cactctggca	ccccaccctc	tctatcccca
109561	agagcctctc	gtctagggac	ccactctagg	atgcagctcc	tttgtcccca	tattgcaggg
109621	tgcctcattg	tctttcacccg	tcattccatg	ggcagccagg	tgactgaaga	agctctcaat
109681	gggtcccaag	caggtccccc	tgtctcttgc	ccaaacttca	ccccagccag	caggtctcca
109741	cattcccaagc	acagatgtca	cagacatgaa	aaagtgcctc	ggacacatgc	ctctctccca
109801	gaagcctgcc	tctgccatgc	tggggacctg	aaaagagggg	gcacgcccca	ccccacgttg
109861	cgatccaac	ccagatgtcc	ccctgagca	gatcacagct	tcacccctga	ccgtgttaat
109921	gtctgcattg	cagggaactg	taacctgtgc	aaaaccggaa	aaagctcttt	gaggaacatt
109981	cccaacagca	tccaagggaa	ggaaagaaga	tatcaattca	ctgtgtctgt	cagatgatgc
110041	gaggcagag	agggaaaggg	gagcgagagc	aaagtctcga	gtgagcagtg	ggcagcgccc
110101	ctgtgcctcg	gctaattcca	gcagccctcg	tgcacttcca	ccagagtaca	tgtggaatgt
110161	cagacccaga	gttctgagat	cctgaggggc	tgggtgagct	gctctaggcc	ccttatgtgag
110221	ccagaaaca	agcatgtcgg	gtggagctga	gcacactgac	ccctgcacca	acagcagcag
110281	gctggggcga	gctgagcgcg	ctgaccctcg	caccagcgtc	gaccctctga	ccaacagcag
110341	catgcggcga	ggggctgagc	acactgaccc	ctgcaccaac	agcagcatgc	cggtgtggag
110401	tgagcacact	gacccttgca	ccaacagcag	catgccgggt	ggggctgagc	acactgaacc
110461	ctgcaccaaa	gtccccgcgt	ccacccccac	tcacccactc	agtgtggaaa	gcactctggg
110521	gagcagccct	aatgcagctc	ctcagaacca	ctggcccggtc	ttccagaaat	gtccctctga
110581	ctcgaagtgg	ggctgcagcc	aaagccaggc	caagcccaac	atcctccogg	tcaggcttgt
110641	cacctgcacca	cggtgtctca	ccaaagatcc	agggccttgc	cagccacatt	tacctgtgtg
110701	gctccacccc	tgaaccagaa	acctatgtcg	gaggggagag	gtggcggggg	aaagggcaagc
110761	ctgagactca	gggactaaat	ttatgtcaga	ggccaggggg	agtttctcgc	cagctgtgggg
110821	ctgtcagcgc	gagtgccagg	ccagagccgc	gtggaggggg	agcacaacct	ggcagctctc
110881	cactgtggaa	taaggggggc	ccagccctcg	ctggggctcg	aaagcctcgt	tctgccttea
110941	cagcccaagg	gggtggaaag	accacagggg	cagcgcctgc	cagcagcagc	ttccagccaga
111001	gagcccgcca	ggccccggag	ctggcccttc	ctgctctccc	cggtctgacc	ggcccaactc
111061	cagctctgga	gaatgtcgag	cttgcgcttt	cccagcgccc	ccccttctcg	acccttctcg
111121	cggtggccac	atcaccgcgc	tggcatctcc	aacagcccat	ccaagggccc	ggctctgacg
111181	agagcaggag	gcgcgcggcg	tctgccaaag	gtgcgtctag	ccctgcagag	cgctccccgg
111241	agccgcctcg	ggtttcacaa	tcagtgtctc	tcctctgggt	ttccggggcc	tggcgtcgca
111301	gctcagctcc	aggggtgaaa	gagaggatga	gctgagagga	ctccaggggt	ccagcccgga
111361	ggcgtgaac	tcagtgtgct	ggcaacagtg	gcttggcagg	gccccagctg	accagggatg
111421	gagagagagc	acctatcaac	agccagagag	ccctggcctc	aggggttctg	ggactgtact
111481	gggaagtgc	cggttggaga	cacagagaca	gttctgggca	caagacgctg	gccaggtcca
111541	gggctcacc	tgcaaatcca	gccaaagccc	tctctccatt	ctccttagg	ccccaaagtt
111601	caacacttag	acatgtcagc	tgaatcatga	cccccgagc	tcattgtctc	aatcccccaa
111661	aaactcgtag	cattacctta	tatggcagaa	gagactttgc	aaatgtgttt	atccagagtt
111721	tcttgagatg	gggaaagtta	cctggatggt	cccgagggcg	ccataaattg	atccagagtt
111781	tcctttctaa	agggagagct	ggctatagaa	gaggaagaag	cacgtgacca	ggagtcagga
111841	gactgcagca	atgttgcacc	aagccaagga	agcccaacag	cccccagagg	ctgggagggg
111901	ccaggaatgg	attctgcctc	ggggcctccc	gaaggaacca	cgctgccag	cactttgact
111961	ttagccctta	aaatcctatt	tcagagtctc	gagctcccga	actgtgagga	atgcaggtat
112021	ctgtgtttag	gccattaaag	tgtgggtaac	ctgtgtacag	agcccaagga	ctgagagtgt
112081	cagctgcaga	gctcacaggc	agaaaaggctc	cttgacctgc	catctacagca	attggcacc
112141	cccaacaaag	cacacggccc	tctgtcccca	tcttgggccc	ccagccccc	acctggggca

112201	agcctcacat	ccgctgtgtc	agtgggggct	gctatctgct	cgccactgaa	tggatgagtc
112261	tcgctgtgct	ccagcttctg	tggtcttgaca	gaacagagct	ccctctctgc	tcctgtgcca
112321	ttgcttcgac	ctcccccaac	agcctccctc	acactggaac	cagtlacgac	ttgctgaaac
112381	actcaactct	ttgtgagact	ggagggcgcc	atccccactt	tggtgcatct	aagctgtgct
112441	cttccatgtc	ggctgcacca	cagctctctg	ctccagcccc	tggtgtctgc	acacccaggg
112501	ggaactgtca	tgcttgaggt	ccggcccaac	ccccgcggcc	caactctccc	ccaggtatccc
112561	ctgcggcctg	gaaactccct	accacaacta	aggcaaaatc	acaggggtgc	gactgtgggt
112621	cgagaggggc	ttcttctggc	ctggtgggga	ggcgtggctt	gtggctgaca	ggtggacggc
112681	acgcttcagc	acatggagga	gctgccaglt	gctctctgct	gctctctgct	tccttgaggg
112741	cccttcattg	tgacgagggg	ctgagagagct	ccccgtgltg	cgltgctcgt	tgctgccgtc
112801	ggagggagct	ggctgcagctg	gcagcagagc	cagagggctg	gcagccgggg	tgctlggggt
112861	ggaaggagcc	aagglcactg	gcacglctct	gtctgagctc	tcctctctca	gcccagcagg
112921	gtgtgtgccc	agcagcctgg	gagcagctcc	ccaggatggg	aagcacctgc	ccctcctgtc
112981	tcgccaggcc	aggccggggc	gagcccgagg	agagcagccc	acagagggag	gcagaagatg
113041	ctctacatct	tgcatccaca	aaagcatggg	tagccaatgg	gggacccggt	tttgtttgat
113101	ctataatgtg	agcacggagg	gggaccccaa	gcatgacagg	lccactggg	tgkgagggc
113161	ccagggctgg	aggcagagag	ctgatgtgga	ggcttgtgtg	gaaagagggg	tttgttgtga
113221	tttgtcgtgt	gagccctgga	ccctggcagg	gtccttgtgc	cagcagcctg	ctctgtgtct
113281	ccaacacagg	acttctgtag	tgactcccca	gacccctgag	gctctgtctc	ttcgtatgct
113341	gataaggggc	gtctctctga	cccggglcgg	ctlgggcccl	cccaagccag	ctgttccga
113401	ctcactgagc	tcagccgcgt	ggctggattc	ctggagcctt	ctcagttcat	ctctctatcc
113461	caaaaatttt	agccacattt	ttaaactcag	gaaagtccac	ataaaaatgt	tatgtctcgg
113521	atttttttga	ttgatgttct	agctgcacca	gaccacaaat	ctacaaaatt	gaggtccaag
113581	gaglttgaga	gggtctgtcc	ctccttgcaq	ggcctcccaa	gcacccctgt	ccaggtgcga
113641	ggccaggggc	ggacacacag	ttgtggctga	ttcccactgc	tgtttctttt	taagtgaagt
113701	gtatttttgg	tttccacgtt	ttcaatcaaa	ggagagaaat	gaaagccagg	caaggggtgg
113761	gtgcatttat	ctttttttcc	ttcaactgct	ttcttgttgg	catgtgctgc	gcagccctgc
113821	aggtcttgga	tttgagctgc	ttgtttttct	gcaaatgggt	ttgtgtgtct	gaacacagga
113881	ccacatgtct	aacacgtgac	tcgggggtct	ttctcactgt	tgccaacact	ccagagatcc
113941	caaggtcagt	ctctgggaag	ctaagacagg	ccctcagltg	cagcagcctg	ttgtcctctc
114001	tgtgcacgtg	gtcttggagt	cagccgctcc	aggtccctta	acggcgggcg	tgggggagga
114061	ctgcagggct	ggagactggca	cttagacatg	cagggaagag	ctcccaacag	acaggggtga
114121	gggttccagt	agggagggcag	caggacaggg	gcttgcctgc	gctgtgcctc	tggtctatat
114181	gcccttggga	ggcgtlaagt	aagcctacag	tgctgtcctc	cccgctgacc	ccaggggtga
114241	ctctccgggt	gtgctgtgca	ctgctgtgcc	agccttgtac	gtgaggggca	ttcacatcct
114301	gtagctgtgc	cggtggtgcc	cccaaaagatg	ctggaatcct	atccctgcaa	cttgagacgg
114361	gtacctttcta	tggaagaaaa	gggctttgca	gatgtgctgt	cagtaagagct	cttgagaatg
114421	gtacccctgc	ctggattact	cggtgagtc	ctaaatgtaa	tcaggcgatc	cttgcagagg
114481	ggaagcacag	ggacacgtca	gaagaggagg	ggggcgacat	atccgggggc	agagattggc
114541	atgatgcagc	cacaaagcca	gggaggctgg	cagcccccaa	gagcggggat	ggggcagaaa
114601	cagattctcc	cctggagcct	ctggaggagg	cacagccctg	ttcaaccttg	attcttggctc
114661	agtgagattc	atttcagact	tttggtccca	agagctgtga	gagaggcaat	tgattatggt
114721	cttttttaca	atgctgataa	agacataccc	gagactgggc	aacttacaaa	agaaagagat
114781	tatttggact	tacagttcca	cggtggctggg	gagggcctca	atcatgtgac	aagatgaag
114841	gcacgtctca	catggcagca	ggcgagagag	agagagcttg	tgacggcgaa	ctccccgttc
114901	taaaaccctc	agatctcggt	agaccccact	accatcatga	gaacacagtg	gaacagatcc
114961	acccccatga	ttcaatcatc	ttcactgggg	ttccctccat	aacacatggg	aatattggga
115021	gtcacaagat	gagattcagg	tggaagacac	gagacaaaac	atatcaattg	ttctattggt
115081	tcagcttcgg	tgaggactcc	ctctccaaca	gtcccaacac	ggatgagatg	ccacacaccc
115141	agagagagct	atggcacagg	gtgcccctgt	ctgtatcccc	agcacagagt	ccctggacac
115201	agagctggcc	ttgggaaggg	cagagaaaag	ggatttgaga	tgctgtcgag	aatgcctggg
115261	ggagacactc	aaccactgct	ccccaaaagg	ccacccagcc	tcagagctgc	gctcccggtc
115321	ctggccagggt	ttccagggcca	cagcagactga	gcaggacccct	tttgcccaag	ccagcatggc
115381	cagggagggag	agaaaaggac	cctccccgct	ccctcacccc	ttccctccct	taataacatc
115441	ctctcccggg	tgttgccata	gctgccaggt	gcacacactc	atccctggcc	tgctctgtga
115501	atgctttaaa	accgtcatct	ttctatcggt	ttccccaggg	aatctgcaat	ccagcagact
115561	atcttggatg	acagagacta	ttagtgaata	tgatttccgt	aggtccatgt	agtcctccag
115621	ttcacccctga	gcagcccggt	ggcttctcgg	ggaaaaacag	acatctctga	aaggaggcag
115681	ccctgacgct	gcagggatatt	cctgccaccc	agacagatgc	ccacccccca	ctgggacaccc
115741	cagggcagct	gtgtcacctt	ccccggcaga	acataaatag	cagggtgctt	ccacacccact
115801	tcagagcgtt	gcttttgggg	aggcagggaag	aaactgcgaag	ccagcacctc	ctccagcgtt
115861	ctgccagagc	cacactgata	tatcagggtt	ccctcagaat	tcactggaag	aatgggtccc
115921	ctatcacacac	acacacatac	acacacaccc	gcacccaggt	ccttgaggtt	ttgtccagag
115981	agcctgtgct	cctcactatg	cctggccaca	ggacacagct	ttatgtatc	tcacagttgc

116041	cactctgggc	cctgaaaagg	cccaggcaga	tggaaaagctg	gacgcctcag	gtccaccgtg
116101	gccctccacc	cagctcccgcc	actcactggg	cctgggtccc	tctcagtcag	tgtgtcgatc
116161	tcagcccagg	gccacgtggg	aacgat tagg	tgcatggcgt	gggaagctagc	tcactgaagt
116221	gcagcagccc	tcatgcaggg	tctgctgaca	ccagacagga	acatcaagca	agggggaatt
116281	ggcccagacc	ccaagtctctg	ggatccagac	ctatagcccc	agcccacgc	tcacatccac
116341	agctcggccc	ccaggcctcg	agctaaggct	ggaagagctc	acctctcctcc	cgacagcccc
116401	ttaccaaaac	cccaccttcc	agcccagctc	cacccctgac	ctggcagctg	ggcccaacca
116461	gcgtctctca	aggccaatgg	ttctctggct	tatctcgaga	ttgaatatcc	tttctgtgtg
116521	tctgaatttc	aagttctcat	taaaagatag	tggtatgggg	gtgggggggtg	aggggaattt
116581	tcgcagaaga	tgccaaacac	cagcagaagc	ccagaaaaat	aggcaaggcc	ggcccaggctg
116641	tggggaaagg	gctgcccagt	tcaggaaagc	aagccacata	tggaaatctgg	gatcagagac
116701	atggcccaag	gacagtgccc	agctcaacct	acctgagcag	aggccccagc	aaccacccgt
116761	aatctctcag	gacagggcag	tgctaatgac	caagcccaca	ccacaaagtg	ccagccccgt
116821	agctctctgg	gacagggggg	ctgcccctgg	gtctggagca	ctcgtcagc	acctagctag
116881	caaggagcaa	agctggggcc	ctgcccctgg	gtctggagca	ctcgtcagc	cccttcacag
116941	ctggccctcc	tagattactc	ctcttctct	ccctgcaaaa	tagtccaagg	gtggggataa
117001	gtcagagctg	agaccacagc	tggaaaaggc	tgggcaagca	gtaccagctg	gcccgctgtg
117061	ctgcagcata	gcagacgctg	attgttcacc	aaggaaacag	gggccaccgg	ttgcactgat
117121	tggtctccag	gtacattggg	cttctgtatg	gaacaattcc	tttgcaaaag	tacttgatcg
117181	cgttttcata	aattagaagt	atgctgccca	ggctgtacc	ttatctttctg	tcagcttaag
117241	aagtcacaga	gacctacagt	aaatgagggc	gggctgctat	ggagtgcagg	gagttacctt
117301	gtaaatacaa	gcacgggggc	tgcccacgac	caggccccctg	gcagctcctg	ttgtgaagctg
117361	caggtgtgac	cgatcctagg	cccacagctc	cagactctcc	taaccccagg	gtctctgaca
117421	gaggtctggc	gtgagctcgg	cagcttccca	ctctccacg	ccagcagcaa	ccacaggccga
117481	gaactcctga	ccacacgggg	gacctcgaga	actatggggc	tgcaagaacc	catcagacac
117541	caacctccgt	ttcaccacaa	tcagggaaca	gagactgggg	gatgtgagga	gctgtggggt
117601	ggcacagtga	ggccagatct	gaggcgacag	tcagtggactg	tcacactggc	cgctctggag
117661	ctctcttccg	gtttctgggc	ccagggcaca	cgtaggttct	gaqtgaaaa	acagacagac
117721	ccatgagctc	gtgcagagac	tgctgaggtc	gctgggagca	cctccagagt	ccttctgagg
117781	ctgcgctccc	cagggatcca	gaaagagact	gcagggctct	tctcctcagt	acagccagtg
117841	aggggttgga	cctaccacag	gcaccttccc	agggaccatg	gcagcacaga	cgaggagagc
117901	cgctgtacgc	ggctctggag	gcctcttccg	agggcacagc	tcagagagcg	aggcactgga
117961	ctcaggccgc	gccgggggac	cccggccccg	tcaccagacac	ttgcttgctt	ttaggcaaca
118021	gaagatcgct	ccttggacag	ccttcatagt	ccaccatgga	ggcccacagc	atgaaggcgc
118081	aggtgtctcg	gctgcttgte	accacatgac	accagagaac	gagggcccag	ccccagcagc
118141	cctgctacgc	cctgcctcca	gccaggcagc	tgggccattt	ctgagccact	ccggggagtg
118201	ctcagctcca	gcagccaagg	tggtctgtgg	agagctgcct	gagttctctg	gatgctcccc
118261	tgccagccag	ggcctcagag	ggctctccac	acagcctgcc	cacctctcag	agcctcaact
118321	gtgtccaggc	cctgtggccc	cccagagagg	ggggagctgg	gtggccggcg	atcacaccaca
118381	tgccacagcg	cttgctcagc	ctcctgcgcc	gccagccccct	tctccaaggg	ttcgatcagc
118441	aactcctccc	cctccatccg	gatcagacca	gcctgcggga	caaaagacaac	aggatcagat
118501	ttccagacca	caaaaagaca	aggaagggga	acagcgagga	gcccccctca	gggagtcagt
118561	cccacctgaa	gacctgcgcc	agcaccacag	ctgtgtctca	gcttgaaacc	tggtgtggcc
118621	ggagggtatg	aacatgtgtg	cccagagggc	gccccccctg	gcccaccctg	tctaggaagc
118681	gggctctgtg	ctgggtctat	tcaccacctc	tacaccaacc	ctgtctgctt	tcagaaaagt
118741	tactgtgttt	ataaagtcac	aacgtctcag	agaagcaggg	agacatccaa	gggcagctgt
118801	gggctctgic	gtgtccccag	gagggcaccc	accaaaagtc	atacatatgg	gccaaagtct
118861	gcccaccacc	ccaggagccc	agagtccaa	acaaaagtc	aaacagccag	ccggggatgt
118921	aatctgcccc	cacacatggt	ccatctgggc	caacaatatt	ttttaaagct	ctgaagctaa
118981	cactcagcca	gatgagattt	catgtcttga	acatggaaat	ccagctttcc	ccggggaatg
119041	agggccctgac	cacagcgggt	tttgatgga	acacatgccc	ctgcacactg	ctccagtccc
119101	tactgcccc	caactgtctc	ctctgctctc	gtgtgcagcc	accagcccag	ctctcagag
119161	caagctgaact	gctgccccgt	accccctagg	ccagccagca	ccctccctga	ctacaaaact
119221	ctcgctgggt	cctcatttcc	gttggaactc	aggtgctctc	atatgctgag	atttcagctc
119281	ttggcctgcc	ccacctccct	agaccgggc	ttgcaaccac	ccatgcttct	ctctgggctg
119341	ggtgcacctc	gcgagccacc	ccagccccctg	aaacccatgc	tgccaactca	ggaatcccca
119401	ttctcccaac	ccaccacccc	atccccagct	cgggccatct	gtaaacctgt	gttccccctt
119461	cgaggcctga	ttcagacccc	ctcaactctc	ctgcccctcc	ctgcccctcc	gcttccctcc
119521	ccctgccata	ccctgcccct	ccctgccata	ccctgcccct	cagcctctga	gcacatcagc
119581	ctgcacccat	cccatgacaa	ctccagctcc	tacactttta	tcttctggtc	catattccat
119641	aaaccaaagt	ctcacactgt	aagtgcacac	tggtgtccag	cttcagagtg	tttctcttcc
119701	atcacaaagt	tatgacccat	gcaggggatt	ttaaaagctt	gaattttgtg	ccaaacttaa
119761	gaacacagag	agttcacata	aacgcagaca	tttccagact	ctcttgaaaa	tcagaaaagc
119821	tgccatttcc	agcaccacat	gacactccag	cttagctggc	ctgtggccggg	gtcccctgtc

119881	cagtttaggag	tgtgccctct	ggtgggcagc	ccccggcacc	tggcatcttg	atgtgcgga
119941	accacacccc	ctccaaagta	acctcatgat	ggcctctgcc	atatcatcat	cttggttgtg
120001	ggggggcccg	gtggcgcatg	gtgctgccct	ctctctccct	gaccccgagt	gctcggttgag
120061	ccccatccctg	ttgctggcgt	ggagcctctg	cgacgcgaga	tgcagtcagt	gccccccagg
120121	gctctgctgg	cgctggccct	tagcctcag	tgtacgagag	gagctctgag	atgctctccg
120181	ggcgactcag	ctggggccct	cgggagagtg	atgtgcccag	caggggctgg	tggaggttga
120241	ggggctggag	gaacacagag	ccagatca	ggcctggctc	cgaaatgga	gagctcgaag
120301	cgaggaggtg	acatctaccc	aaacccacc	aatcccttg	tgctggccag	cgaaagagtc
120361	agagacagag	ccagtgctcc	atatggggga	aaccgcacat	gccatggcgc	gggaggttga
120421	cgagacagag	ctgtggggaa	gagctgacca	ggtgggtgga	caggagcatc	aaggggcgg
120481	tgcccggtgg	cccatctact	agggggctgt	tgaaggagat	tgttaggcag	gacgggcttc
120541	tggggggagc	cttgatctga	tccagltca	gaagcatggt	ggctgcggcg	ggggggaatg
120601	gctgcggaac	aaaaatgggt	ctggggtctc	ggggagaagg	tgttacaggc	acccgggaga
120661	gggtaaatg	tccagacggg	agcatgagag	acaaggacga	gtggcagta	tggagagagt
120721	gaggagcgaa	atggacaggg	gtcgtgtagg	tagagcgccg	tgagatggga	gggaaggaga
120781	agcactcgtg	ggccttccag	gtcgtggctt	gaagaaacgg	cccccaagt	taccacagcc
120841	cccatccctg	gcagctgtga	gggtagtgg	catggtggaa	agagactctg	cggggtgtgat
120901	caggctcagct	ctgggctctg	ctggggccctg	aatgcagctc	tggccactt	ccctatggcg
120961	gggagtggag	ggggattagg	cagatgcgca	gaggacgagg	ccatggggcag	acaggagaga
121021	tgcagatgcg	cagcttttga	gattgagcgc	atgtggccct	aagccgaagg	atcgtgtcag
121081	cccccaggag	ctggcagagg	ccagaaagtg	tccctcccca	gagcttccag	agggagcgca
121141	gcctctccga	tgcctggatt	tgggtcaggt	tataccgaet	tcggagcttc	gccttcaaga
121201	ctgtgagaga	acgaattcct	gctgttttaa	gtggttccct	acagcagccg	cagaaagaat
121261	attcaggttc	ctggtgggtg	gcgggcgaag	tgggtgccat	tctccgaggg	cgatgagaag
121321	aagggcaagg	ggcgagagac	cctctgggct	caggcatcct	gggcttctgg	ggaagagagt
121381	taggagcgca	tgataaccaq	agttaccagg	gcaagaccga	ggtcgcagaa	gcaqgtccag
121441	agggctgtg	gaggagagcg	gaagaggggg	tgggaaccct	gcagggccca	cgcttctgac
121501	ctccctctg	gtctcaggcg	tcaaccaggc	ctggcgagaa	ctcagcagat	gctctgtgaa
121561	tgagcgaaac	gatgacccca	aagctgcagg	ccaagctccc	caaggggaggc	accgagcagg
121621	caggcgagaa	gaacacagcc	aggcctgtcc	actggcacag	ggcagatgtg	tttaccagtg
121681	caggagttgc	gagggttccag	cactgatgtg	cccgggtcgt	gggttgagaa	agggctgggc
121741	ctcaaatcca	tctttatgac	gaactcatt	ggtatgaagt	caggaattaa	cccttggggc
121801	cggcagcggg	caagctgtcc	tggagcgtga	agttctgaet	tccctggacc	ctttctatct
121861	ggccaggtga	cctcaggcaa	ggtgtcaagc	tccccaagcg	tccaagtctg	ctctgttaca
121921	taggcccgct	catttccctg	gcacactcac	acaggagcac	atgcagcctg	gggaacactg
121981	agaaacggcg	cagcctccca	cccaagtgat	ttcatccagc	caqcttgcct	tccagagcag
122041	gaacacggag	cctagatggg	cagacaggct	gctggggaca	cacaggttct	agggatcaga
122101	actcacacct	acacggaaa	ggcatctccc	agacaggctc	ctggcgagac	ctggggctta
122161	tgagccaatt	agggcaggac	actggggcat	gaggaccctc	tatggcagct	ccatcttctg
122221	ggggaaacttg	cagcccaagt	cgggagccat	cgatgccaca	ggcctgaagc	ccctgggtta
122281	ttctggaggg	acctagtgtg	acaatgaggc	caggtctcagc	ctgggggaag	gcagtaacct
122341	ctggcgcgag	taaccggggg	gacagtttgg	ggtctccctg	accagcaggg	gctgcgctcc
122401	ctccctccgtg	tgccctccctg	aggacaatca	ctgggctgag	cccaacgac	tagctccatc
122461	tgatgtgctc	ctgggcttgc	cctcctgaa	gggcaaccct	gcaggcgact	ctccggagac
122521	caacagagaca	cactggggtc	catctgaagc	agggcctggg	ctcaggtctc	ctccacagcc
122581	cgacagacctc	caagcgggtg	cattctccaa	caagacgttc	caacatgtct	ctccacagcc
122641	cgcccgcttc	tgttgaccac	cgtgcgatgc	gcacaacgca	tgtggagcga	tgaacattgt
122701	ggcaactgoc	ccgagaccaa	aggtcgacc	ccccgagacc	aaagctgac	cccccccgca
122761	gaccacaaag	tgaccccccg	cgagacccaa	ggctgaccgc	ccccgagacc	aaaggtgtgac
122821	cccccccccc	gagacccaaa	gctgaccccc	cacccgagac	caaagctgac	cccccccgca
122881	agacacaaag	ctgaccccc	cgagacccaa	ggctgacccc	ccccgagac	caaagctgac
122941	ccccccaccc	gagacccaaa	gctgacaccc	ccccgagacc	aaagctgac	cccccccgca
123001	ccaaaggtgtg	acacccccca	gagacccaaa	gctgacaccc	ccgagaccaa	aggtctaccc
123061	cccccgagga	ccaaaggtgtg	aacccccccg	agacacaaag	gctgacaccc	cgagacaccc
123121	aggtctgaccg	cccccgagac	caaaggtgtg	cccccccgcg	agacacaaag	ctgacaccccc
123181	cgagagacaa	aggtctgacc	cccccgagac	caaaggtgtg	cccccccgca	gacacaaagc
123241	ctgaccccc	cccccgagac	aaaggtgtg	cccccccgcg	agacacaaag	ctgacaccccc
123301	cgagagacaa	aaaggtgtg	cccccgagac	caaaggtgtg	cccccccgca	acacaaagct
123361	gacccgcccc	gagacacaaa	gctgacaccc	ccccgagacca	aaaggtgtg	cccccgagac
123421	ccaaaggtgtg	cccccccgca	gacacaaagc	tgaccccccc	gacacacatc	agctgtacgtc
123481	ctccgaacag	cgctggcgag	aatatgctc	ggggcggggg	gcactctcag	gaagaggtgg
123541	agatcatttt	attccctgat	tttcaaacag	ttcagtatat	gcatcccgag	gatctctgtg
123601	ctaaacacaa	cagaagaaaa	ttgacaggag	agcaagacac	aaatgtgag	ctatgttttt
123661	gctcattttt	tccgggact	ttgatattcc	tgtctcagtg	gtcaagtgtg	tctgtttttt

123721	catgtaattt	ttattaatta	gaatcagttt	cctctcccca	ctcaataaat	atataagaaa
123781	ctccagagag	acacttaggt	gaaaagcact	cagttgggag	cctgctggtg	gcctctctgt
123841	ccaaccccac	agctgtccag	cctgggggtc	tacagaaaag	ttggtacccc	gacctggctc
123901	taagtgtgag	tcccgtctgt	ctcttgaagt	ctgtggatgt	ctgggaaacc	gacccaggtt
123961	ctccctgtgg	ggacagagag	aggatcatgg	ggagggctct	gtccctgacc	ttcaggtgtc
14021	acccgagggg	ccagagagcct	ggtagtacc	gggtcatcaa	gtcaagtgtc	cggctattcag
14081	atccctggtga	gttcatcaaac	gcctgcctct	cctcttctct	ctatagaacc	tgctctcttg
14141	cagttggatt	gttggttaaa	tagttactag	aaatggcaac	tagatgttaa	aaattttata
14201	agttccaaag	gcttaacacg	tctcctctat	tactctgaat	ttttaagtat	tcaaccacag
14261	acagcctcga	agttaggggt	tccacggaaa	cacaacgatg	catcctgttt	ccttctgcag
14321	agcagccccc	ctggctgggt	cccactgccc	cttcccatgg	agggccacag	tgccagcttc
14381	agcagccccc	caactgtgta	tgactgggag	ctggctctgt	actcaacccct	gtcaacccctc
14441	agcataaaca	ctgtcagggt	tggtccaaag	ttgcagtggt	aaacatcaag	catagaaagg
14501	acaaagcata	ggaaagcaca	aggtccatgc	acagaggggc	ccagactctg	ccaagggaca
14561	aggtctgcaa	gtctgtgtgc	cctcaaacct	gatgcctctc	tgaataacaa	caacagcagc
14621	accttgccaa	gagtccecca	gatgtggcca	ctgtctaaag	gtccactttc	tgaagatatt
14681	ctaactccaa	gattctttct	cccactcttc	cttttggtaa	tgaacctggg	agatctctgt
14741	gacaaagggg	caatggccag	gcagcctccc	ctgtctlgac	agctgcctgg	tggaaagacg
14801	gtctccacag	gagagctgca	cagtgctgtg	agttccccag	ggcgtctctc	tcagagcctg
14861	ctccggcact	gtctgtcctc	ctccagctcc	ccaaggaggt	ctataaacac	cttgggagct
14921	ttcaggggag	ggggctggag	ccctcactac	tgaacaggtc	caggagtggt	aaagtgtcat
14981	agttgggttt	tgagccactc	cgctctacc	actggacagt	cccaggagtg	ggaagtatgt
15041	agtgagattt	gagccactgc	gtcctcacca	ctgggggctc	ctttacagtt	cagcatccca
15101	gagattttca	cacactgaat	tttccacttt	tccaaaaggg	agccacaggg	caggagtggt
15161	tccatggggt	taggaaacag	ctgtcctcac	ccgcactctc	accagctcct	cctctgtctc
15221	atgggattcct	gcacaggtgc	cagccactcc	cgctctctct	ctttgagagc	gacccctttc
15281	ctaggctctc	gtcttggagc	tatggctgga	gtggggcctt	ccagggttct	gtcgagcaca
15341	gtgaccacaa	cagagtttgc	cttgagctcc	acagaaactc	atgactttgc	ttggccactc
15401	tgaacctccc	tggaactggt	ttcaggccct	acagaccttt	acttccagaa	gtctcttttt
15461	tgcccttaag	tgccgggagc	caatttccaa	caggaccacg	gagcccccag	agctacatgc
15521	ccagcgagcg	tctaaggaaa	gctagactcc	actttgaagg	gcagggtggg	ggcgctagga
15581	ggcaggagga	gtaagcactg	gacgctgagc	caggggccct	ggtttccaat	gggtgcatgc
15641	cgactctctgc	cccttgcagc	gccccagtt	ccctcatctg	cccaactgag	agctggacta
15701	cgtaacctat	gaattactgc	ttatctgttt	tattcatcac	cccaaaaagg	aaccccgcag
15761	acgtttggcag	gtaccgatca	tccttctaaa	actctactag	gaccacagag	aaagacacaa
15821	accaaaacc	agaaaccac	tcctgttaaaa	taaaaggccaa	ggccgtccca	tgccacgcga
15881	ggccttccat	gggtgccagc	cagccctctg	tcctctgggc	cacccccagt	gtcttctctc
15941	ctccctgggc	ttctatttgc	acctcaaaag	ctacacacac	gccttctctc	tttgttatag
16001	aggtcagtaa	gtacagagaa	gctggtaaag	atctgtcagc	accaaactga	gactccccct
16061	gactcaacaa	gagagtttaa	aatgcacact	aaaagggaat	catttttcat	tataaagttc
16121	aatgcttttt	ttctcttttt	tttgagacag	agttctcgtc	tgctacccag	ctgggagtg
16181	atgtggacaa	cctcagctca	ctgcacactc	tgccctccat	ttttaagcga	gtctctgccc
16241	tcagcctccc	aagtaactgg	gattactggt	acacacacc	acatccagct	aatttttgtta
16301	tttctagtag	agatggggtt	tcacatggtt	ggccaggctg	gtctcaactg	cctgaccctc
16361	gggtgatccac	ctgcctcgcc	ctcacaaagt	gtggggatta	caggccaata	ccatggccgc
16421	cgccagcttc	aatgcttttt	aataataaac	agctttatgg	tgctgtgatt	cccttcccat
16481	acatctcagc	catctaaact	atatgtgtaa	agtgaaacat	tcaaccataa	ccactattga
16541	ttttagaaca	ttttcatcac	cccaaaaaga	aaccocgtac	acttggcag	tactccccca
16601	tgttttcccc	aacctctcct	ccccaaccca	gcctcgggca	accactaatc	tacattctgt
16661	ctctatgaa	tggtctattc	tggaacattc	atatacgtg	aatccatgag	tattgtggct
16721	ttttgagctg	gcttcttcca	tttagtatgt	tttcaaggtt	catccatggt	gtagcatata
16781	cgagtatatt	attcctcttt	tattgtcaaa	taatacccca	tcacatgatt	ataccacatt
16841	ttatttatcc	atttgccggt	tgatggacat	gtgggttgtt	gggtttttgg	tccattgcat
16901	aatgctgcta	tgaacatctg	tggaacaagt	tttgtgtgga	catgtgtgtt	ctactctctc
16961	aaagtggaa	tactgggtca	tacagtaact	gtgccagact	gtcctctaca	ggggctggac
17021	ctactctcag	gccccagca	gtgtgtgagg	gttcaatttc	cctcatcccc	acccaactat
17081	cattattctt	aggtctcttt	actctagcta	tcacagcagt	gtgaagtgtg	gtcccatcat
17141	ctctcgctgt	gcattttccg	gacacactaa	gatgcataag	atcttttcat	gagcttatag
17201	gtcatttttt	taactttgga	gaaacatcca	ttcagatatt	ttcccttatt	tttaataagc
17261	cattttcttt	ttattgacct	gtaagagctc	tttatatat	ctagacacaa	atcccttatt
17321	agatatataa	tttatgaata	ttttctctgt	ttctgtgggc	tgctttttca	acttttctat
17381	tagtgtccac	tgaagctcat	aaggactttc	ttgaagttca	gttcaacctt	ttttttcttt
17441	tgctccactgt	tgcttctggt	gtcatatcta	agactttgoc	taatccaaag	taacaaagat
17501	ttactctcac	attattttaag	agttctatag	tttagctctc	tacatgata	tactgatcc

127561 atccattttac agttaaatttt tgtgtatggt gtgaggaagg ggtccaactt cttttttcca
127621 tatgttgata tccaggtgtg ccaacacact ttgttgaaaa gaccatttct tcccaccactg
127681 acttgctctg gacctttgtc atgaggaagt atttccaaac tctcagttct tctcagttct
127741 attctaccac tctatattgtc tatccttctg ccaacacacc acagttctga tctactacagc
127801 tttgtagtaa gtccttagaat tacaaaagcat gagccctcc tctgttcttc tctccaaatg
127861 catcttggct agtcctgaat cctgttattt ctgtatgaat ttttaggtaca gctgttcaat
127921 ttctggccca agacttagctg ggaatttgata gaattaaatg catctctggaag tcatgtttgg
127981 taccattatg ttcaactcct agggcaatgt ttaggaaaaa ttggccacaa aagaacactt
128041 gcatctaga ccaagggctt ggtcaggctc ggtgctcac accctgtaat cccaagctct
128101 tggagggctg aggcaggcgg atcacagggt caggagattg agaccaatctt ggctaacacg
128161 gtacaaacat gtcactctga gatgggagcc ctaagctcat caactaacct cctgagccac
128281 tgtgagcaat gtcgtggcccl ggagagaggt gcacagaggg atggcttcaa cctgtgctgt
128281 agagaaatgc tctggagagc tctgtctcag tctgtcttgg catccctccc gacgaccccc
128341 cactgcccac tttacgcagc ctggacttga ctcttaaaat gagacacctg ccaagtgggg
128401 aaatccacgg tcgaaatagc ggtgctgccc tctgcacta tgcctacaggt ataattaggg
128461 ttcaaatca gaagacatt agcatattt ctggaaaaaa aaagctatgt aatatataaa
128521 ctttactata tttacttta ataaatgggg gcaagcaaaa aatgtccaaa acgccccaca
128581 acaatcacag aaagagtaga ggaagaaaaa aggaattaca gtttaattgc acrtggttcc
128641 taactctcag aattcttcag agtactgctt agtttcttaa aatcagtaga aatcagtaga
128701 tacaagctga agaatctcat tcaaaattag aaaaagtcac actagtaaaa ctaagtata
128761 cctctcaaat gactgaagc agaccatata gaaaatatg tagactagaa aaaaagcaaa
128821 aaaaaaaaaa aaaaaaaaaa tccagaaaaa cagaaaaaat caaattagat cagaagtaa
128881 gaacaaatat ataacataaa atgttaggctg ggcattgatg gatcccaagg cttctggagg
128941 ctgagggcgt taggaatttg agaccagctt ggcacacatg gtcacacccc aagttagcga
129001 aaaaatacaa attagccaag cgtgtgtgtg cacacttga atcccaagta ctcaggagga
129061 tggggcggga gaactcactg aaccaggag cagaggttac agtgagccga agtgactga
129121 ctgactctca gctcgggaga cagagacaga cctgactca aaaaaaccaa aacctataaa
129181 taacaaaaca aacaaaaca ataaatggct agattattat tattattatt attttggaga
129241 ctccagcctg ggagacagag agataccctg tctcaaaaaa acccaaaatc caacccggcg
129301 cagtggctga gctcctaatt cccagcactt tggggagagc aggcagattg atccaaagc
129361 cagaggttca agccacagc gtcccaatat gtgaaacccc gtctctacta aaaaacccaa
129421 aattagcccg gcttgggtgt acgcgctgt ggttccagct gtttggagga ctgagggcag
129481 agaatgtgct gaacccagga ggctgaggtt cagtgagcc gagattgcat cactgacac
129541 catcttgccc aacagagtg gaactcatct caaaacaaa caaacacaaa aaaaacagc
129601 acaaaaacaa taatgtgtca gattattatt attattatt ttgagacaag gtcttctctc
129661 ctacccagc ctggagtgga gtacagtggt ctaactgag ttaactgag cttctgctc
129721 ccaggttcaa gcaattctcc tgccctcagc tcccaagtag ctgggattac aggttgcaacg
129781 cagagcctcg gctaattttt tgcaatttta gttagacgg ggtttccgca agttggccag
129841 gctggtctgc aactcctgac ctacagtgat ccgccacact cggccctccc aagtgctggg
129901 attacaggtg tgaaccacgg tccacgcca gtcagattat tcaataaaa cttgtgccc
129961 tgcaaaaaaa ttttcaaaaa ttacagatt tatgttaggc taaaagagtt tctcagatt
130021 tttgaccact taaggaacag acacacccat gacattttg tgaccgaaga agggcgcaa
130081 cccagctgca gcatgacctt cttacgtaaa atcatgtata tctctgtgac aaagctgcaa
130141 acagtctaat aattaagcat gtggcctcgg agtcaactgt cctgagttca aattgttact
130201 tcaactctga ctgagtgccc tctctgtgga cagtgtctg agggttgaat ggagataaat
130261 acaggggtttc catggctgat gtgagggcta agcgagatac cgcgtctcta gactggcca
130321 cagcagctga taatcagcc gactggcagc accatcacca tcatgtagt catggtatga
130381 cagctgttag gaacccgcta acaaaagggtc agcaggttta cgtcgtgggg ttggttttca
130441 tttttattaa tattttctgt attttttttg ttttaccatt ttacaaaagg catatagcaa
130501 aacacagaat tcatgtgaag acagagagga cggaaaaggg gagagccacg agagaaagat
130561 aagggggaaa atggaataaa caagatgccg acaagaatgc ctacccgggg cagccaggg
130621 gccagccagg cggccttctg ctaaatcaag caacgcttga gtggctccct cgggacctgt
130681 tctgtccaga atgtccaca ctacgcccct tgacggcag tcccccatct ccatgtctcc
130741 cgcacacacc aatgccgggg tagcctccag gaacacagc tgccagacca gcccacga
130801 ggacacagag tgcagtgggt cagaccaga ctctgggact gcaacagcgg tgctattttg
130861 tgcagagaa gtacactcgt cccggccaga tctctggct gaactttccc agtctctgt
130921 cagctgaag cccacgtcgg ggaacgtccc cagtcacaaa ttaacattca cataaaac
130981 agggcgtctc tactataaga agaaaaatga ctgctgtgc aggaacattt ccaaaactat
131041 caaaaagcaa gatttttctt tctgagcttc catgtctcac cagctggaac acagagatt
131101 tggccttctga cagctgtggt ggagagagc tcatgtaata aggtccaaat caccatttca
131161 agactctgcc aacagggcgg gcacgggtgg tcacgctcat aatcttagca cttggggag
131221 ccagagcagg cggatcacct gaggtcagga gtttgagacc atctgggca acatgtgaa
131281 accctgtctc taaaaaata caacaaata cggggacatg gtggtggggc tctgagtcac
131341 cagctactcg ggaggtcga gcatgagaat cgtttagacc cagaagggcg aggttgcac

131401	gagctaagat	gcgaccgctg	cactccagct	tgggctacag	agtgagagac	tctgtgtctca
131461	aaaaaataaa	aaaaaataaa	agaccctgtc	gacagggccct	tccctgggct	ggcaccctgcc
131521	cacaccagct	ggaacctaac	acctttgac	agctgttccc	tggccagctc	ctctgcctcct
131581	tggaggctctg	cacacaccgc	ccctcccttc	ctccaaccag	cccgaggagac	cagcagcgct
131641	gagcaccagc	actcccagag	acagagccgg	caagggttgc	ctgtctgtctc	agggggcctgg
131701	gggtggcagg	gcacatgtct	ggggagttgc	aacatagtct	tccctaccag	atcccccagg
131761	gggaaccacc	tggagagcaa	ccaaacccct	caaggctgca	ctccaaagac	tggaggccctg
131821	gggaccaccg	acgctcccgc	agcatctcaa	cgaggctcac	ccacatccac	caggagaaagc
131881	tgcggcagtg	gactgcaccg	accccttccc	acaaaaggta	gggggcggcg	ctctccctag
131941	gtgtctcggg	ctacacatca	agggcagcac	agggcggggc	tgaatagca	gaacctattc
132001	tctcaacagt	ctggaggctg	gagctctaaa	atlaaggltg	gggcagagggc	ggtttctctc
132061	gagggctctc	tcttggcctt	gtcgacagca	tcttcccctc	gctctcacat	ggcgctcccc
132121	ctgtgtgtac	tctgtgctta	atcgcttctt	cttagggcac	cagctaaica	tgttgaaatc
132181	gtgacttcat	tttcagttaa	ttatctcttt	aaagacctg	cctccaataa	cagccaatat
132241	ctgatctac	gggagtttag	gacctcaaca	aaggaaactg	gggatctagg	ggggcgctca
132301	ctcagccaat	aacagcctcc	agggcgcaag	tccacacaga	atcgagagag	gctgctggca
132361	gggtccgggt	gccagagctg	ggcgggctgt	tcccttggga	gcactctggg	acagggcgga
132421	aggtcaatgc	caggatcaat	ccccaggggc	cagaaatcca	gcatggagga	cagctggcac
132481	tccagggaaq	taggagcccg	ccccagggct	gccacatggg	atgcctctct	acacctctct
132541	tctaaactgg	gcattctcac	atggggagac	ggggcagtga	agggcagggc	aggagacaga
132601	gaatgtctga	cagccggtag	gtggtgtgag	ggcctcaaga	ccccctggcc	acagcagggc
132661	cactctctca	cggtccagca	ggagccctgc	tccagagccc	agactgagtg	ccagctggcg
132721	ctgtgtgtct	acattctctg	agagagggct	gctaaggctc	ccgtgtgggg	gcactctccc
132781	tgcacctgcg	accccagaat	ttctgtggcg	atcccaggc	tgaagtctag	cagctgttca
132841	aagacggggt	atgcaattct	agatgtgtgg	cgtgttttga	acagcatcca	ataaccttct
132901	tgtggcagct	cagcccgagg	agcaggtccc	cagctctcag	gcccagagag	caggggttgc
132961	ctccttccca	cagcgggtgc	tcttccagga	ggggccaggt	gctcagaggg	aaacccagtg
133021	ccaccaacgg	ccagctgccc	tggctggctc	ccagcagttg	ggggaagaca	cactaccgca
133081	cgagcgtccc	ccttagaagg	agggggcagc	aggaactccc	atggggagag	cacagcgctg
133141	gacgcagccc	gcccaacgce	agtcccccca	ctccactgtg	caacaagagt	cagcctttcc
133201	tctctcaaaq	cctcgggttcc	cccagagggg	gctgcagttg	agagtgtgat	gtgaagctca
133261	tgtggggcatg	tcttccatgt	ctgcttgggc	aaagctgccc	atagagacca	tttttccagc
133321	ccagaggaacc	ccccctttca	ctttccatca	caggctgtca	tctccccggg	tccctggggc
133381	acaggccccct	ctcaggaccc	ctccggcttc	atgcacacac	agagcctctg	cactgctcca
133441	ctctctgcttc	ctttgaattgt	ggcttctttt	ctctgtgtga	tctctggggc	ccactgttgt
133501	atccagggca	cgctgaccgc	cgaccgttct	ctgcccgcct	ggccagctgc	cactctccgg
133561	ggggcagccc	tgcaggcccc	gaggtgtgtc	gcttccccct	ctgtccagag	ttcccaaaaq
133621	aaaaaatcgg	ccataccate	tgcaggaagg	tctgtctcgg	ttttatttatg	ggaagagcca
133681	gggctctgag	ggccagggttg	tgtgggctgt	ggccaggctg	ctctgcaggc	cttcagagcc
133741	tgcaccagcc	gtgtccccc	ccaccagggg	tggctccggg	ggaggggcgg	gggcaggacc
133801	cagccagctgt	tgaagccaca	ctggtagctc	tggagtgtgt	atttatctca	gagccagagca
133861	agcagccgce	agggccctag	gcccctctgt	gggtggttgg	cccactggag	agccggagac
133921	gcgagccagc	caatgaccac	agggcccgta	tccagaatgc	agggcagctc	agggccagct
133981	gccttccctgt	gatgccacta	cagattccgg	gagactgaac	tcatgggac	aagtgcccca
134041	cagatctctgt	tctcacaag	ggacaaggcc	accaggtgtc	acactgggag	cagactggcac
134101	cccagggtat	taaaagggcac	accaggctta	tgtccagtaa	ccttgcctgc	acagcagcagc
134161	acgggcagcg	atccataccg	cagcacagcc	ccgcctccgc	cgagttgcac	gcccgggagt
134221	tacgggcccc	caccagaaac	cagacaggtg	agcaggtggc	agctccaggc	cagctctcga
134281	ccttgtgact	accaggccct	ccccgagagc	aggtctcagc	cccccaagct	gttctagctg
134341	ccacaggact	gcgcctccct	tgcagggtct	tgtcacaacc	cacactccct	ctgcagagac
134401	ggcctgtccc	accgccccgt	gtgaaatcac	acccaagcac	agcggaacga	tggagagggc
134461	agctggaagg	tgctttcggt	ggtaaatct	ggggcaatt	gaacacaata	ataaatgtat
134521	tgataagttt	agaaccatt	gaataaata	aatatccaca	gatcagttat	ggtagaatac
134581	aataactgaa	taataaataa	gtgggagggg	aagggcagct	cttctccaca	ggcttaattca
134641	actaaaaatg	taaaagaaat	gatgagatag	aaaatcatgg	tagtaaaacc	catcgccaac
134701	agcacagctg	ggaactcggc	aggtacgact	tatccgtgga	ggctacaatt	actggagata
134761	aggtacaaca	gaacagaaat	gctcacatca	cgtaaccaga	gctaccacag	ggagatttaa
134821	tcatcgcaat	gggaaaaaag	agtgaagag	tggttgtctc	atcaagccac	ctacttgaga
134881	cgtgaaaaaa	cctcacagtc	acagccctag	agctcccggg	gcactcaggt	gaatccaggt
134941	aggttaaccc	ttagccaaat	gaagaagggc	atcactgaca	ataggagaca	cactcgacatc
135001	ccaatgtgac	agtcacagtc	tggtgcccta	cggagggcac	agtgctcttc	tgcagactctc
135061	ccgagaaaca	cgtaagctca	agccagttaca	agggcaccca	gaaaaaccca	tggggatgct
135121	ccgtctacaa	aacagctgac	cagttcctct	caaaaacatc	caggtctatga	aaactcagat
135181	gggaggaaca	gggagagaca	caacaactaa	atgccacgtg	ggatcctgag	ggggagcctg

135241 aaccagacaa cggacatgag tcgggaagact gacaaaattc agtgcaagtct ggaagtctac
135301 ggttagaatt tgaactgggtt gtccaccaca aaactcatct tgacatttgg tccccagtgt
135361 ggcagtggtg gaaggtgggg cctgggggtga ggttatctgaa tcatgtttgg cgatccagtgt
135421 tgaatggctt tactactgtc tcaagtgagc aagggaaatt ttgctctggg gaagcttctgt
135481 tagtcccaga gctctgggttg ttataaaagc aggcacactcc ctgggttttc ctctctttga
135541 cacatgggct tcccttttga ccccttccac cactttatga cacagacaaa aaagggctca
135601 ccagaagcca gggccatgct cttgaacttc ccagctctga gaaccatgag ctaaaatagc
135661 tctctttctt tataaattac ccagctctcag gtattccatt atagacaac aaataagact
135721 aaaaataaga cactaagtac caacgcttag tatcaacct acaagtctct tctctcagtc
135781 cttgttctac tgttcataca gaggaaagctg ggtgaagggc acagaggaaa gcttatgctat
135841 ttttgcactt tttcttgcaa ttttccaaag acagcataga tctaaagtta tttccaaatt
135901 aaaaacttaa acacataaaa ataaaaactgg gccccagcct cctctcagtc agtctggccg
135961 gtccattctt cactttggct gcccttccct ggagtgaggc gcgccactgc atcccagtg
136021 tccaggaggc aggtcctgta gaaggtggta ggttccagg atccgggagc aagtcacagc
136081 gccctgagaa ggaagcaagc ctggcggtgt ccagggaac ccaggagaa cgtgtagcgc
136141 gacaaatgc cttgtgcaga gggcctatgc ctgagccgc ctgctgtccc tgcctgtccc
136201 tccaggcaac tctccaacc tctgccctgt gcatggggcc acgattgccc agagagctct
136261 agggctgttg actatgaggt tgcctcaggt ctcagtgaqt gcttgatga gcgagccacc
136321 ctactctgcc agggaggggac ttccttcaaca ggtgtgaqt ggaactgacc accagcatga
136381 tctgtacagc tgggttttga ggaattccca tcacaacaga cagaatgagc tggggctctg
136441 gggataacac tgcccaagcc tcaagctga gtggccactg tgggctctgt actcggctgt
136501 ggggccaact ccgggctctg gtgtccagg gtgggagctg tggggcagca tcttcaact
136561 tctgcctctg tcaactgggt ccaggggcgt ggcccaacc ccaaaagttt cagtgtctat
136621 tggggcgagt tctccagggc tggctgcact cactggcttg ggtctccaag ccgtctggga
136681 tgttgcctgc caatgccagc acccttcccg tgtctctct tgcctccaag ccgtctggga
136741 catcactgct tctcccaagg ctgggtctga gagggtatga atgctctga atcctgggag
136801 ccccttctct ttgtcattct gcctgtggtt ggagagtggt ggaagggcgg agtgtgccca
136861 cacttaaacg aggttgaatt ctgggagtg cagatgaaca gcttcagagc tacagtggag
136921 gatgaggtct caagatgtgt ggttcttccc tctcgtataa aggttccaaac gatattttt
136981 tatttattta ttattttaga gacagatct tactgtgtca cccactgcag tacagtggct
137041 caaacacagc tcaactgcagc ctggagccccc tgggctcaag tgatctcccc aactcagcct
137101 cccaagttag tgggaccaca gacgtgtgcc acaacagcta gctaattaaa agaaaaaacct
137161 aaattttagt aaatgaggtc ttgctatgtt gccagggctg atgttgaact cctgggctca
137221 agtaattccta ctgcttgggt ctcccaaaat gctgggatta cagatgtgag ccagccacc
137281 tggcccaaac attaaattta acacgaccag atgaggacgt atggctcaca ttttcttgca
137341 cactctgggc aagatccaca ttggtgtgtca tcatataga agcacgtgaa ccaccttcaa
137401 gggaaacacg gaagccggca tctccggggg cagagagaag cggcgagatc agggaaaatg
137461 ccaacacacg cactaaatc atggggctgg gaaatgcggg gaaagaagt atacagatc
137521 tccacacaaa caggttccaa agagccacac tagctgggac agcacactc accataaaga
137581 ccccgctgt gggagatata aatatgaaat tagctgggac agcacactc accataaaga
137641 acaggaaagc atagctcctg agcctgagca gtaacagggg tctctgcaag cactgagcag
137701 acagtgaaga aatgtgttgc tgccttctct cccctgcctc taactctctc catgacatc
137761 tctgatttcc ccagtgtggc tcatgaatct gtgtataaat tttttaatca aattgtctat
137821 atagacagca gctctgcagg aaatatatgc ccaggggccc atgtactcta agggacgcca
137881 gccatgctt tgggaacagc ctagaagagc tgtcaggaag ggccctgggt ggggtgggag
137941 cagcttgcgc aaggtgact tctctcctg caagatacat agctgagaga agcccaattg
138001 aacgagggcc aatccaaact cccaccacac cctgagtgcc agaccactgc ccagacggct
138061 tctgaggtc ctgctgggg ccttccaaag ctacgtgag ttgaggtcag ttttccatc
138121 ccccaatagc ctgcttgaa ttgttctctg cttaagttag ttgaggtcag ttttccatc
138181 caaagagaag cctcctgtt tctctagagg ttgtcaggag ctacgtgag cgttccatc
138241 ggcttgggac atcctctgtt gttgggggga gagggtggga agcatcttct ctgcagcact
138301 gacctcacct tggtagacag cacagcttct cactcacagc atctggtgaa ctgcataact
138361 gaagccgctc ctcttgttc acgggggaaga caagtctgacc ccaatttct ctgctcccg
138421 aaaaataaga acgttatagg tgttgccttg tctctctct gaggaaaagt agataaata
138481 gctaattctc acaaaagagg cagcgccctg tgctctctt gctcctctt tggataatc
138541 gctgactccc acatgagcag cctctgatct ttgacagct ttgttatgga tgcgaatac
138601 ggcttgaagt gaaaactgac atctcaatgt gaagagtatt tctcctgcta actgcagtg
138661 gtcagagcgt tgcctgagc tagacagata gggaaactgg ccttcccatg tgcctgag
138721 tgttgcacag cctccgggac acagcccaag ccaatagatc acagccatgg gacatcagcc
138781 aggccagaac aggcagtgac ccaatagatc cagcccaag cagccacagc gagtggcat
138841 tccagggtag ggttttgggt gtttggttgt ttttctcaag atggaactct gctctgtcag
138901 ctggagctgga gtgcagtggc atgatctggg ctccaccaca cctccactct cgtgggtcaa
138961 gtgattctcg tgtctcagcc tcccgagtag ctgtgattac aggcacacac caccacaccc
139021 agctaaaatt tttttttttt tgtatttttt tagtaaatg ggttttcatc acattggcta

139081	gggtggtctcg	aactcctgac	cctgtgatcc	gcttgcctcg	gcttcccaaa	atgctgagat
139141	tacaggtgttg	agccaccgca	cctggcccaag	gtttttttgtt	ttttttttgtt	tttttttttga
139201	aggtatttgtt	gcactctcaaa	caatccactt	aatgacttgtt	aaatgggggcc	tgcattccca
139261	gcacttttttg	gagggccgagg	caggaaagact	gctttagcccc	aggagtctcca	gaccagcatca
139321	gtgaggccctc	atctctacaaa	ataatttaaa	aatttagctgg	gtgtggtggt	gcatcgctgtg
139381	ggtccagactc	acttgggagg	ctgaggtggg	agaatcaccct	aaacccctgaa	ggctccaaagt
139441	cgagtgaagt	aagtgctgct	ctgactgactc	cagcctgggt	aacagaaaaaa	gacactttca
139501	aaagaaaaaa	atatgactgct	gaaccgggagg	aaagatgttg	atgaaggagcc	gagcgcttca
139561	ctcagagctg	acagcccaaga	acaacgtcca	gccagccaat	tatttttaa	gcattgccct
139621	ccttttggfga	ctcaatccca	tgaacaggga	agatgggtgc	atggctagaa	gctctgtcca
139681	ggctccgagg	gctgcccagc	agaccacccc	taggggtgct	caggatcaac	ccgaccaggg
139741	gatgctgagc	accatgtcaa	gtctgaccca	ccacagtgcc	ccaaggaggc	agcttgttga
139801	ggctccggaat	gggtggtctg	agattgggca	cccaagact	tggggagttg	aaggggccacc
139861	ccaaggccctg	cccccgcccc	cactgcaagt	cccccgagg	gctgcaacca	catccgagg
139921	ggctgccccg	cccggtgtgc	ccagcttcca	aggtggcctg	ccctgatgc	cagaaaaact
139981	tcttgcctgg	caagcgagg	ggctgctgt	ctagggtcag	caggaaatgt	caacattttc
140041	cctgaaaaat	ccttcaattc	cctcattgag	ctccagtgat	acttagtcca	tgtatacact
140101	caaacaggaca	actcttttcc	accccaaatg	ctgacagcct	ctgagcaggt	ctaaagagac
140161	aaggaaagagt	caactgtgca	actcctggca	tggaggccct	cctgtgcctc	tcaactactg
140221	cctctaggcgt	aacgggctct	tggggaaact	gggggggttc	ctgctggcgc	ctactgactt
140281	gtttccactc	caaaaaccag	ccttgagcac	cctactggag	gctgctgggt	tgtctgtcca
140341	ctcagagcgt	ctgagagcat	ctggcacca	gctgtcccta	ataaatagat	tctggagttag
140401	aaagaaaaac	tggctttgaa	aagacacaga	gaagaaaccg	gtccctgctg	cccttctctc
140461	caggctgatg	tcttggctct	gcgggttctc	gttgggtctc	ccaccccttc	cctctgtgct
140521	ctgtctctaca	gccagtagg	caagactgat	ggactggctc	ctcctggctc	tccctgggtg
140581	aaggggggcca	gccacaggaa	gccagggtcg	gggagagaga	cagaaggttt	tatttctccc
140641	gctgtggttg	ccccagcct	ccccggcctg	caagctccct	gctctcgccc	cagctctgtg
140701	gagtgtgtgc	tttgtcaact	cctttctgtt	gatcccttat	gtgtccaggc	cccttgactg
140761	actccagcat	ccctgtagcc	cgacacagga	gccctcactc	tcatattcca	gtgctgtgtc
140821	ccattctcag	ggtttgggat	gggaattcca	caactccagc	caatttccca	ccctgaagac
140881	acgggtgacc	cagttccctg	caaccaccca	agactcagga	cacaaaattcc	caggtctggtc
140941	cgaaacatct	tgaacagggc	agaggagccc	cgctcatggc	gctggggcca	gtcccgagac
141001	tgaagtgctg	gggtggaggga	gctggaggct	ctggaaaggga	tccaggagca	gccaccaagg
141061	ccaacttcca	cacacgtctc	caaaagccct	tccattttcc	gtcgaaccca	ctagtggggc
141121	ctgtgctggg	agcaaggagg	cgctggagac	aggtggggct	gtggaaaggag	gctgccaatg
141181	ggagggaagg	ctctccaac	ccaggacgcc	ccactcgagg	aggaatatgcc	tttctgtggt
141241	tgttccagga	gaactgaggt	cctgctcagc	cacacagcac	gtgccaaagt	gcagagccca
141301	gggggaaggc	aaaccgagga	gctggcggtg	cctcgagcga	gtcagttagt	gtgaagcagg
141361	aggaatgggg	agggctgagc	cgctggggaca	ggagaatgac	ccacaggtcc	acacctctgt
141421	ggctccgtgg	acctcagcag	acgcttgggg	tgatgccaga	ggagccatgca	ggaggtgtcg
141481	cgaggggctg	tgcttctgag	aatgctgaca	atgatgggac	ccgggaagcac	ccggagctcca
141541	agggagcggc	aggggcaatg	ccggggcctg	ctgtgctcgc	gaggaggggg	agttgggtga
141601	ccaagagagt	gggtgagcag	ggagcgggaa	ggttcccaagg	ggggatgtgct	cctcgagagg
141661	ccagcagctg	gaaatgtcga	agtagatcct	tcagcagggc	tggggccctc	gtcgagaggc
141721	aaagaaagtaa	gggagggagg	atggggctgg	accacagcaat	tcagcaagga	aagcagaggg
141781	cggagagcag	tggtcacagg	agggagcga	agggagcga	aggacaggtg	cgactccagc
141841	cggataagga	agggaccgca	ggcaggtgga	gatgggctca	gtgggtlaagg	tccccagggt
141901	ggccttgagg	agaggcgagg	agggagaggc	ggcgggcaga	gagagctgtc	tgcggtgtgt
141961	gggtcaggag	ggtggtgcct	gggaagacca	gctgcctgcc	tgctccccag	gttgacaggg
142021	gcaagctcaa	gcataaagat	gcaaaagttg	agggccgggc	aaggacgggc	acttggtatc
142081	actaacacaa	ccatcaatg	gtttaacatt	gcagttagag	atcaaccaga	ctgagatgga
142141	acaaacagag	catcgctctc	gtaacacatg	cctgaacatg	gctaaaagag	agcacagaaa
142201	aactgcacaa	ccgaaagaga	gggtgtctgt	agtatcatca	aaatagaact	cagagcgaat
142261	actaaatgac	tgaaaactcc	tggtgtatgt	acaaggcttg	gtgaaagcac	acagctacca
142321	agctttacaa	accaaggggc	gcagcatctg	ccatggttag	aaacacaaat	gctgtcaacg
142381	ccttgttcaa	gggaatggc	acggcccaat	ccagtggtg	gagagctctc	ctgacagaga
142441	cgagaggggc	acagactcat	tgactcggga	aatgggcccg	attgatttct	ggatggaatc
142501	gagactactac	aaccgaaagc	aaatagatgt	tcttttcaag	catgggggag	ctatttcaaa
142561	aatgtactccc	ggcctccctc	gcctgctgtg	cctgaatccc	ccaagcctgt	gagccgttag
142621	ctccaccacc	acctgctgtg	caatgcagaa	acctgagatt	gctgtctctc	ggacacagga
142681	accaagaggt	gaagtgtgta	tggggacccc	cttctgcccc	actcctgtct	agtgactgtc
142741	agccagaagg	cctcccccac	agcctgcccc	ttctctttga	ctgttccctt	cccocaatcg
142801	ctgaggcccc	atgtctccca	ggacccctcc	tcaagacccc	ggcttccctc	tccagggtga
142861	cttggggccct	tgcataatg	ggcataatgc	caactcccca	ccctactgta	tgaagtact

142921 ggcagagatgc ccccaacgccc cctccactata atccagaggg accggagagct ctgtctctga
142981 gccagagacaa cccactgcgcc accaccagct cctccagggc tctccacaag cccctctctcg
143041 cctcccctgc tgcgtctctcg gaactctctct cctccaaagt ctctctcagc tccctctcagc
143101 aacccgggtct ccttaactctc gaacacagaag actctgctct cctccccagc acccagagcctc
143161 cctctcccctca cagccctctct cagcagggtcg gectctgctc gctcccccac cgcacaaatg
143221 cccatggcctc ctgtccctccc cccccagggc cccggggagc tcccaacacct cctgctcagc
143281 cctctctctgg ccgccccccag ctgctcccggt tctccagaga cccagagggcca gggctcccctc
143341 gtggccacagt gaccctgcmaa cgcagatctcc gtgctccctt gaaagggcca cctgacatct
143401 ag3461 gacaaacacc aggagacaga ggcaatttca gaattcccgc cctgactggg
143461 gaaocccagc tctcccctct acctcaatcca tgtggcagc cgggectctg caagcctcca
143521 tggagatgag ctctctctgt caaagggtga tgggtggaag cgggectctg aagggtgatg
143581 aggttcagtag ggtgggggacc cgttlaatgg gatagtgag ctatccagag agggccaggg
143641 agcttgtgga cctccactgc tgtgcgaaca cagcaagaag gccccagcta tgagccagga
143701 agcagccctc accagacacc aaaocccact ggcgccagt ctcaactgtg agactcca aa
143761 actgtgagaa ataaatttct gttgtttcta aacccccagt ctgtgatatt tgttatggc
143821 agcctgaacg ggggacacc tgcagagcct gggatgctca gaatacacac gttggcaggg
143881 ggggtggag gggctctaac gaactgggag ttggctagaa atgtatagag ggttgagagg
143941 gaggacttca caagcccttg taggaactg agatctgccc tgggttaggg tgggtgtgca
144001 ctggaaatgc aagatgcaga tataaatgct gacaaagacc aatttaaac cagagatgct
144061 aaggtctgaa acattttgga ttactactagg aatcaacgca aagcttaata acctaaatc
144121 caaccactgc gaagtttttt caacattctct ctaaatagct ctaagttaac tggaaatcaa
144181 atcataaact tttttaaaat gatgacagtg aqaattcagc acatcagaac ctgtggaga
144241 gcaacaaagcc accaccagaa gcaagttcac aaccttctta atatagacc agagttaaaa
144301 taagtgaagc cagtgctcca gtcaagagg gcaagagag agcaaccata aacctaaaga
144361 tgttgataca aatagcatga gtgacaagat ctcaactctg agtgtgacat atattgtaaa
144421 agatggaaaa accggttagtt tgtatagctg taacaaat ta ctgcaaatgt agtggtctc
144481 aaacacaaatt ttgtccagcc gcggtggctc accgctgtaa tcccagcat tggggggccc
144541 aaggtgggtg aatcacagag tccaggagtc gagaccaat cagctaacac ggtgaaacc
144601 ctactctactc aaaaatacaa aaaattagcc aggcgtgggt cggggcgccc gtatgccag
144661 ctactctgga cgtgtgagga ggagaatggt gtgacccag gaggcgagc tgtcagtgag
144721 caaagatcac accacagcac tccagctgag gcaacagatc aagaactcaa ctcaacacac
144781 cacacacaca cacacacaca cacacacaca cattttattc ctgacagctt
144841 tagagatctt aagttccaca tgggtctcac tgggctaaaa tcaaggtgtg gggagggctg
144901 cattccttct ggaagctcta aggagaatgt ttcccttgtt ttttccagc tttagaggtc
144961 acccaactct cttggcttgt gggccctctc tccatttcta aagccagcac cggagcatct
145021 actgacccct cctgactct gaacttctgc tccactcttc tacttttatg aatcctgtc
145081 atggccagat cacacctgta atccagcac tttgggagc tggagccggc ggatccaag
145141 gtccagagat ggagaccatc ctggctaaca cagtgaaacc cctgtcttac taaaaataca
145201 aaaaattagc caggcgtggt ggcaggcgcc tgtagtccca gctactctgg aggcctgagc
145261 agggaaatca cttggaccca ggaggtggag cttgcagtga atggcacttt atggcacttt
145321 actccagcct gggcgacaga gtgagctccc gtctcaaaaa aaaaaaaaaa aaaaaaaaga
145381 gaatccttgc cattacactg tggcactccag gataatctcc ttactttaaa agtcaactaa
145441 tttagcaactc gacttccatc tghtaacctc actgcccctg ctatgtccac tgattttatc
145501 ccagggtctc gggactggga catggacaac tttgggagaa ggaattattc gctctccaca
145561 gagactactc caaaagggtga ggaatttcat gtttttgcct tctccttaat acctttctac
145621 attattcaat gataatgacc atatgttttg atgtttttg agtaattttt aatgtgaaat
145681 aatgtaaatt attgtgaatc atgtttttgc agtgtttgtg aataaattgt aatgttaaac
145741 aatggaaaac taagaataca aaagaataag atcctcttat agcaacctat gatgtctttg
145801 caacactttt tggctgggtct ttgtccccc tccccctgac tttgagctgc ctgagacaa
145861 gccagcccc ccttcaactc tgtggctcta gcaatagca ttgagacttt ggccagcag
145921 gctttagcag acattttcca aagtgaagat aaatccccta ataaaggag aagctcgaa
145981 agaagaatga tgtgcacact caaaagtcga gatctcgaa agcttcgaa taacacagat
146041 gatattattc tatttaagag tcttcataca tacaacttaa ttcagtaaca cttctgagaa
146101 ccgaagccag cagacagata accgcttatg cctctgctc tggcaaaaa tcttcactca
146161 ccggggagcga gaagaaagcc aatataaagt ccggagaggtc cccagagaca gaggccggag
146221 gatgtcccca cacactgagc acaagcccca aacacaggcc tcccagactt ggcgcggcg
146281 gccagccccc tcttggcctc agtgaagctc ggaactctca cctgttccaa gtggggagaa
146341 ctggtccagc ctgcaagaac cgctactctg gatctgggca aaggttctgg aggcctctg
146401 ccgctcctgc cccacagaaa aaaaaaaaaa aaaaagcagc agagggcatg cctcaggggc
146461 cagcctggc tgtcatccct gaacatcctg ggcctgaagg agaatatgag aagactttcg
146521 acttgagactc ttttagactta aatccccact ctgtcacttg tggctatttg gctctttcg
146581 accctctgctc tccggtgtct aaaaatggcac aatattacct actttacaga tttgtcaagg
146641 gattgaaatc aagataatga agcgaaggga atgtgaccta gtaactattc gccatttta
146701 cagatttggga aacaggctcc aagagatgac atgactcttt caggattaga gaccataatga

146761	gtgtgtggtatc	cggtatctgtc	cgcacagggc	tgggtctctc	cccagccccc	caactgggaca
146821	ggggcatctc	ttgaggggtg	ccctgtggcg	cccgatgtgc	gcagggagac	aaggtctgtat
146881	gcacagatct	ctctccactg	ggcagcacct	gagttaactg	gggagagaag	tgggtttcaa
146941	gcattctctct	ttgtaacacc	tgggtttaag	tatgtcgaac	ttcagttttg	acgcagaact
147001	tggagacac	ccactctctg	ggaaaagaca	caaatgaggg	caggaccctc	gcaccatgtg
147061	ccctctctgcc	ccagcagtg	ccctctatcc	agcgtgtgaa	ggaggagtcg	agccggagag
147121	ccggtggcag	caggcgctgc	gggatgcccc	gctcacagtg	tgtctgtgct	gggagaggag
147181	cgagggcagg	cccccagctg	cttgggaaca	gtcagccctg	acacctcccc	ctctgggtgcc
147241	cttccaagag	agccctcagg	ccccaggctc	cgggactccc	gtcagaggaa	cctgagatcca
147301	gtctgggtccc	agccacagccc	ttccggcagg	ctctggccag	cgctcaggaa	gggagtgtag
147361	agttagtagt	agccacacaca	ggagcagaag	ccgggacttc	tgcctctcac	cccaatgcac
147421	gtatgaaat	agcgggggtg	gatggccggg	tgcggtgggt	catlgtcgtg	ctcccgatct
147481	tttgggagtc	tgaggtgggc	ggatcatgag	gtcaggagag	ggagaccctc	atggcttaaca
147541	cagtgaaacc	ccgtctctac	taaaaaata	gtccggcggtg	gtcgagagag	gtgggggggg
147601	ccgttagtcc	cagctgtctct	ggcgacagag	gcaagaccgt	ctcaaaaaaa	aaaaaaaaaa
147661	aaaaaaaaaa	aagcggggga	gtgataaagg	ccccgacaac	ccaaactcga	ccaagattct
147721	ctcggcaggc	agagctgctt	ccctctggga	gtcctctgga	ccaggaaagg	gctctgaggt
147781	cttgtggccc	agtgggtccg	gcttcccaac	ctccactgtg	ccagcagcac	cgatagatgt
147841	ggccacaaag	gtgcattcac	atttgcgacc	tgtgtgtgaa	tgcaagacct	gggtgtctatc
147901	tcccgggggg	gcgaatggca	gtaagcaagc	aaagagaaag	tcgacagcac	gtgacgcagc
147961	cgtagagaca	acagcacaga	gtattgttga	ggggggcggg	ggacttctgt	gaagagatga
148021	cactgtagct	gagccttaaa	tgtattggaa	gagcgacgca	tggagagatc	tggtgaaatg
148081	ggcttcacga	cagagacaag	caaatgcaga	ggacccgagg	tcagaaacga	cgctccatgt
148141	ttgaggaaga	ggaagaaggc	caatgtaggt	ggagtgtgga	gagcaaaagg	tggcgggggg
148201	cgaaggagac	gagggcagag	tgggtgggct	gggtggggta	caggagagta	gagtggttgg
148261	cgggggcgca	ctgcacagag	cttctctctg	tggggggagg	agattgagtt	ttctctctgt
148321	tgcagtgagg	agccccctga	gggtgggaag	gaccllgatt	ccctgtcagt	tgtctctggc
148381	attgttggag	aattgctggg	ggcaggaatt	ccctccctct	ctctatgaaa	ttctctgtta
148441	gcagggatct	gcctatggct	gtattttctg	tgttagaagt	ccctccctga	ttataaacgc
148501	ctcaaacatc	agagaggttc	ctagggaaag	gttcaggttc	ccagtagacc	gagctcaatg
148561	gcctcacctc	catccccacc	tgcctgcagc	caccocaaat	cactggattc	atgagatgct
148621	agctgagcac	ctgcacaaat	aaataaacag	atgcagattg	gggaagagaa	ctagaaatgt
148681	aggtttcaga	aaccgcgaga	gtttactctt	catctctctg	tattctctac	ctagatataca
148741	agggcaggtc	aactccagaa	cggggcaact	ggatcacaga	gcagagctgc	gagacaagac
148801	ctgttgttgt	agctcaagga	tagaaaaagc	atttccgaat	gctcagagcg	aagccaacac
148861	ctcatttttc	tttctatttt	ccctttttct	ttctctgtgt	gttccacagc	gagctcaatg
148921	gcagccacag	gcacctaaaa	ctctaaagag	ggagcctggc	tgtgtagctgt	tgaatgtctg
148981	ggctctaaag	aggtggagtg	aaacccctat	gctctttctc	cttttggctc	cccagggagt
149041	gtagtgtgca	gaagtgtgca	agacagcagg	gtcaataaag	ccccagcttt	ctggccagag
149101	agattttaaca	agagcccgca	ggaaacttga	cagggttagg	aagatggcag	agacagcaga
149161	ctgtggggga	ctggcatcag	aaagtgtggt	agaaaactct	gggctctgtc	ctgagccgtg
149221	cacagtgagg	ctcgacctca	acaaacacgc	cataggtttc	ggggactgac	ttaggagatg
149281	gaccaatgcc	caggtccccc	acgtgccact	gggttgtgtt	tacatgggat	tgatcagaat
149341	aggaactgac	aggttttgaa	aatggacctg	acattgaacc	gcaatccaca	ggacatgtat
149401	aactttgcag	cggaacccaa	taaacccaa	ggtttaattac	ctgatcaaac	aaaaataatg
149461	agccttccca	tagcatttaa	acaagatcca	aagtctcata	gggtaataac	taagatttca
149521	ggatcacagt	caaaaattact	tggcatacca	agaacccaga	aaatctcaac	ctctccagcg
149581	aactacaacc	aacagataac	aagactgaga	tagcacatat	gttggaaatt	actgacattg
149641	atcttaagac	agctattata	aaaatactcc	ataaagtaag	ggcaaaaaat	ctggcaacta
149701	atggaaaagt	agttcagca	agaaacggaa	gacatatttt	taaaaaggac	caaacggaaa
149761	ttttagaact	gaaaaataca	ataacagaaa	ctttaagaac	tcaactggat	ggctcaagag
149821	cagaatggat	atgacagagg	aatgagtcag	tgaacttgag	tgactagcaa	taagacttac
149881	ccaattcaaa	ctgcagcgag	taaaaatatt	ttataatgcc	aagagccctt	gaagactgtg
149941	gaaaatccag	aaagatctaa	ctaacagcta	tgtcccagaa	ggagaaagaa	gaagagatga
150001	tactaaacaa	atatttggtg	aaataatggc	tgaaaacctc	tcaaatctct	tgaaatgcac
150061	aaacactacca	attcaacaga	ctggcaaaccc	tcaaatgaga	gaaatccag	gaacacata
150121	ccccacagtt	ataaacacac	cttgaaaact	atagacaaaa	aaaagtcttg	aaagcagctc
150181	aaagaaaaaa	gtgcattact	tataggggaa	aaatgggttc	cttcaagagt	gatttcttat
150241	caaaactgtc	aagtccaaag	gaagtgcaaa	acattgttaa	aatgctgaaa	aaaaaaaaaa
150301	agaaactgca	accagaattt	ctatctccag	caaaaaata	cttcaagagt	taaggagaaa
150361	taaagacatt	ctcagatgaa	gcaaaaacta	tagaattgtg	ggccagcaga	ctaccttaaa
150421	gtggttgtgt	gaagggaagt	cttcagacag	ggggaaaagt	atgctagaaa	gaatttttga
150481	acatcaggaa	taaaggaaga	gcaacagaaa	tacctgaaaa	aataatacac	attagctctc
150541	tgcgtatttt	aaagaggaat	gataatgctg	tctgagtga	tttctcatgt	atgtaggtgt

150601	ataaaagaca	acaacattca	ggggaggaca	gagggaccta	tatgatggca	gggtctcaac
150661	attcaacttc	aagtggtaaa	atatgtattt	gaaaaagact	ctgaagagtt	acatatgggtt
150721	attataatcc	ctggagcaac	cagtaaacac	tatacaaga	tatacaacca	aacacaatag
150781	gtaaattaaa	attgaatacc	aaaacatgct	caaatagaaa	aagcaaaaata	agcccaaaagc
150841	aagtagaagg	gaggaatgat	taagagttaa	aatcaatgaa	actgaaaaata	tgaaaaataat
150901	agagaagact	aatgaaaacca	aatttggctc	tttgtaaaga	tcagctcaact	gataaatcttc
150961	tagtgagact	gaagaagaga	accagaaaaa	aatatgaaat	accagtatca	gaatatatttt
151021	aaaaaggata	tcaactaatga	catcacagcc	ataaaaaaga	taatgtaggt	gtataaaaaa
151081	ttctatgcatt	ataaatccat	cagtggtgaa	gaatgcatg	atttctcaata	aacactataag
151141	ctcccaaaac	ttaccaaaaga	caaaagtagat	agcataaaaa	gttctgtaac	tattaaagaa
151201	aaagaattca	tagtttaaaa	agctgcaaaa	aagaaatttc	caggtctaga	cggttttact
151261	ggaaaaattc	actaacctct	caggaaaaaa	tggcaccaat	ttctcccatc	attttatgag
151321	gtcaacatta	ccctgatatt	aaaaccagac	aagacagtt	caataaaaaa	aaaaacccta
151381	caacaccaata	ccctcatgat	acctataaac	caatatccct	caatgaacat	agacacaaaa
151441	atcctcagca	taaacactggc	aagctcagtc	cagccaatga	taaacacgag	agtgaccaac
151501	agacacata	tgtttgtcct	ggatttgcag	agtcctggcac	aatatccaaa	agtcagtgta
151561	acaacagcta	ttaaacagact	gaagaaggaa	aaaaaatatg	cttatatacaa	gtgatagaga
151621	aaaccatatg	ataatgattca	ttcatgacta	aaaattaaaa	actctcaact	aaacagagaaa
151681	gaacagtgtc	caacctggta	agaatatcta	caaaaaaact	aaagcttaaca	caatcacoga
151741	ttctgagaaa	ctgttttccc	ctaaataaag	gagcatcttc	acagcttcta	ctcaaaccttg
151801	ggttgagctc	tagacagcac	ctctgtacac	accagcaacc	cagcacagac	ctctcaggcca
151861	gggcagagaa	ggcccacaga	gagcttcccc	aaaaaccogt	ttctctggga	agtggaatgt
151921	ccaggaaggc	tgactggccc	ctaaagacct	ggggagccgg	gggcccctgc	gctccagcct
151981	catgaagctt	tgatctccag	tgcccagccc	gggcccagga	gagagccaga	cagatgaact
152041	caacacaggga	aaggacaccc	aaggaaagag	caactccatt	agatcgagtc	tgacagcaac
152101	aagtgctccc	acaggctctg	aggaacatgc	ttctattcca	ggacacagcc	tgacaggagc
152161	acagcctgag	ggatgagagg	gcaggcgagt	gttctgctgg	gtgggaattt	ggccacaggc
152221	ccagctccca	catctgttgt	ggcctctggca	gtatcaaccc	tgagctgagc	ttctgtgagc
152281	tgaggccagc	acgtagttaga	ggggcagccg	agtaaaagaa	atctctcggg	gaagtccagt
152341	gcagccagcc	ctaaqcccgc	cccagacgct	ccctctgcta	gcatagaca	gcacagagcc
152401	ccgcgcactg	cagagctgcc	ccggcccacg	cccccctgct	ctgcttgact	ccgcacatct
152461	cccacagcgt	cgctgagag	catcgcaacc	tgcaagcctc	gagcagaa	gagcagaa
152521	ccctggggccc	acaggaggtc	caaccggccc	ctcagggctc	gagcaccggc	ccaaaccctg
152581	gtggccagct	ctcagcactc	caggagacgt	ccatgacaca	gaagcagctc	tgcccatgag
152641	cattcaacag	cccggtccgg	ggggcaccac	cacgcagggg	cttcccgagc	catgctgcac
152701	catccactaa	tatgaccatc	acgtatgcca	ggctgggagc	ccgggctctc	agaagctctc
152761	acttcatctc	ctatgctcct	caaggctcgc	ccacagaact	ttctgaaacc	cactgacagc
152821	gtctacgtcc	tgcccaaagc	cctccagcgg	cttcccacca	catgactccc	cggctgtccc
152881	ctgatttctc	cttcaactca	gtgccagccg	tgagagtctc	ctctgtgttc	ttgtgcacat
152941	tcaactcagc	acctgcctca	gggcctttgc	acctgtgctc	gcctccacct	ggaaatggct
153001	ttcttcagaaa	gtccaggttc	tagctccgat	gcctctctac	aaggacctgt	gtgaagggtg
153061	ttcctctctc	atcacactcc	acactgcoat	ctgtctctat	ttttttgtaa	accattttcc
153121	catcatcagca	cacgtggggt	ctttattggg	tcaactctct	ccccctgacc	agggatctca
153181	cgagagcaaa	gttgtccttg	ctgggtccac	gtccaccagac	cagtgctcgt	catctcaccg
153241	gcattgagtt	gagtagcgga	ggctgcggcg	ctccctcagg	ctcctctccc	ctactctcgg
153301	attggtgtta	gtttccttgc	caaggtccct	gtgggtcaga	gtgggccaag	ttttccgctc
153361	ctctggagcta	caacaaggag	aaaacagggg	tttaagtggg	ggatccgggt	gtcttatttc
153421	ttcccagctc	gtaaagcatt	cagctgtctc	gggtggaggt	ggaagacagg	cctgtagagtt
153481	tgctctgacac	aattgttcta	gcagcaggag	acgctgatgg	gatgtcggga	gtcaccaggga
153541	caactctgaca	aatgaccggc	ttggggatcc	gtgagctca	cggtgtccca	gagagagagc
153601	tcgaaatgcc	agctggaaa	gtcacgcagc	ccagctggaa	gggagccggc	cgagcccctg
153661	ggcagctggc	ggacaggaa	gggcagccgc	cactgttcac	gagcaagcgt	gcagggggca
153721	gcagacccac	aggagctcct	gcaatgcgat	gacctagctc	gcagggcccc	ataatagaca
153781	gcctcaattt	ttttttctag	gcaaaacttt	gggagagctc	tttttaattg	tatttgcaaa
153841	acagtttctc	ccaacatgat	gcccattgtg	atgtggattg	ggggaaatct	ccaaagccctc
153901	tcagagaagc	gcctccttgc	taaggtgggt	ttccagcaga	tgaaatctgc	tgccgctctc
153961	agtggaaatc	ggggcggggc	tgccatggaa	ccccaccctc	ctgacaggtc	ctggcccttc
154021	ttcctctctc	ttcttccacc	cccaggggcc	ctgggatgca	ctggaggttc	ctgggagcca
154081	gaacagccat	gaagccctcg	gggtactgtt	ccctgtcttc	agcatcccc	tcagctgccc
154141	attcattcag	ccattcttct	aggaggacag	gccagcccat	ggtgagggcc	gtggtccagg
154201	gaagagatcg	aagaggggct	agggcgccgt	tcttgccttc	gggtctcttc	cgctgcctcg
154261	gcagcccgct	gcctcctgac	tgtggactgg	agttctcgac	ctgcgcacat	ccacagagga
154321	tacatggaga	gatggggcag	ctcagcagag	agggcgaac	accaattcag	ttgaaaactc
154381	tgttttcccc	gggtagaggc	cagttaggag	ccagcatgcc	tcccaagggt	tgtgcaggga

154441	gacttcagcc	agggccccac	tggggcgagc	ccagtcceac	tgtctctgcc	ttcagtgccc
154501	acagggctgy	ggccatttgg	ggatagagct	ggagaaagcg	ccacctctcg	ggccccaggt
154561	cacagatgt	ccccccaaag	ccatcccagt	ccagatgtctg	cagacaacctg	gtccctctggc
154621	aacagctgacc	cagacgtgcg	cttcaactct	aaactgtgac	agctctctag	ccccggcccc
154741	aagcctgaag	gcccgtccct	ccagacaactg	cagaaagctg	gggtgactg	taatccccag
154801	acttltgggg	gcccaggttg	ggcgggtcac	gaggtcagga	aatagagacc	atctgtggta
154861	acatgggtgaa	accctgtctg	tactaaaaat	acaaaaaaa	tgcgcacgca	tgggtggcagc
154921	accggtgac	ccccgactct	caggagcatt	agggcaggaa	atcgcttgaa	ctgtggagggc
154981	ggagtgtgac	tgaagcagata	tcattggccat	gactccagc	ctggggcaaca	cagcaagact
155041	cagcttccaa	aaaaaaaata	aaaaaaggctc	tggccctggc	tgccctcgcc	acttgaccgg
155101	tgtgggtcaac	tctccgtgcc	tgccttctct	ctatttaaaa	tggatagggg	aggagctctg
155161	aggggcaaac	acccttgggt	tgcctcaggg	cagggcagga	cccttggagcc	agcagccccc
155221	cactgcaatg	ccctcctgtt	ctccctggcc	ccaccttggc	ctctgggttc	tggggttcaa
155281	gttcagggag	ctctgggagt	agcctgcggc	ctgggggaca	ccccctgaat	ccccctgaat
155341	ctctgggtgat	ctgcgcaggt	tgatgctctg	actgctctt	ctctgtaccc	aggacctctc
155401	agggagagcc	accctggact	acgcttgccc	agggctccct	gcagctccca	tgtccccaca
155461	gctgcaggg	gctgagatgc	tctgtgtctc	gctgtggggg	ctgtggagag	gctctctccc
155521	ccccacaggc	tgcagctccc	agcgggcagg	gatctaggga	tgtcttggct	atcccccaag
155581	cagacaatac	agtgaagaaa	gacagagaa	tggcttctgt	ctctttgggg	gctcaccccc
155641	ccagctccaa	gtgtgaggat	caggagggcat	ctacaagctc	ttccgaagcc	cccccatagg
155701	ctgctctgca	gggtacagta	tggccttagag	agccccagag	ggctccccctg	gtcatataat
155761	acttcaggat	ttccaggaac	aaggagaggt	tcaagtccac	agaaagaatc	gtattggcgt
155821	gggtggccac	ccaccggtgc	tagaaatccc	gctggggtga	gggagtgtgg	tgcagagcct
155881	ggcttatgcaa	aattgtggct	ggagcagcct	ggggggccccc	gggagcttgc	tagagatgtg
155941	gaatccaggc	ccccccagaa	ccagctgaac	taaaaacctg	attttggcaa	gttccccagt
156001	gggtctatga	cattgaagtc	tgggacacac	tgctctagac	ctctccaggg	gtctctaaag
156061	gttgggttaga	agggcctact	gcccctgccct	ggggacgag	agggactcag	gcttagtccc
156121	ttcttgggag	aggtgaaagg	ccaccacccc	agaccatggc	caaatcgagc	gggttagagg
156181	ggcatgtggc	agggacagcag	ctcagcaggg	tgccccaccc	caacctcccc	tgctctcccc
156241	taaccacttg	ggcgctgggt	ctcttggaac	ggcccccct	tcacccccca	gttggagcca
156301	aaagacaccc	tttccactcc	ccagcccccc	accccccaac	ccacgagatg	aaagggccgt
156361	tgtactctat	gtcaacaaca	acaggaatacc	atttttccag	acagcatat	cttgaagctc
156421	gatgcgtatg	ttcagcagat	gttcaactctg	agggcgagcc	ttccccccgc	cttgaacgct
156481	aagggcacgt	ccccatccct	ctcccccagg	gctgtgcgct	cttggcccaag	gttgagctgg
156541	ccctctgggt	gccaactctc	tggaaggggc	ctgtccagct	ctgtctctgt	aggggctcatg
156601	ccccctcact	ctgtctctct	tagcccatgg	gcaggcgccg	cttggctcag	agggcacaag
156661	atggggcagg	ctcagacagt	ggtgagcagc	atgccaaaga	tcacaggggac	atccacctgc
156721	cagcaaaagg	tcaatgacaa	caaaggacct	gaacagtggg	cactgacagg	gaactctctg
156781	cttgaccacaa	ctttagtcca	gctcctgaac	cttctccttc	gccccacttc	ttgttaagct
156841	cagtttataca	agaaactctgc	taagtccagca	atgccccate	ttctccatct	tgatcacctc
156901	tgtatcttgc	tcaggctctg	ctctcccatc	acccctctgg	tgatgtctgg	tcaccccagg
156961	ctgtctctag	ctagaatcat	gttaggttag	tttagctaga	atcccccgac	ctctgattcg
157021	ctcttagtaga	ttccccatcc	cttgccccct	acccctctcc	ttggctatac	ccccccactc
157081	ccccgtctgt	tattcggagt	tgagcccaag	ctctctcccc	gactgcaaga	gcccactctc
157141	gtatcatgta	atcatttttc	atttaacaac	actctgctct	cttgaaggga	aggttagaat
157201	gacagcagtg	gggagggagg	agaaggggca	tgggaaacct	gtgcgcgtaa	gggcatactc
157261	tgaacctctg	gccccagctc	aattgttagaa	gctgcagagg	tgcaaatata	acagtcagac
157321	gacctagcca	gtggctcagct	gaaaataagg	ccccgaagac	acgtccgggt	ccccatcccc
157381	gaaacccata	aattgtacct	tccatggcaa	aaaaagaact	tgtagatcgc	gttagtccct
157441	ggatcctgag	atggagtgc	tgctggattt	tcgggggtgg	ccaggagata	gacagaaatg
157501	ttatattccc	cccaaatccc	gatgttgaag	gctactcccc	caaggtgata	tattaggagat
157561	tggtggccttt	gggaggtgat	tatgtcacca	gattccatcc	acacgaatgg	tatttctgcc
157621	ccgatccagag	ggaacccaga	gagctccctc	accccttcag	ctgttgaggga	tgcaatagac
157681	tgccctctctg	aaccaggggag	cagacctcca	ccacaacctat	atgacagct	tcacaaacgc
157741	tgagaaatga	gcttccgatg	cttttaagcc	gctccgttta	tggtatttgg	ctatagcagg
157801	ctgcatggac	taagacaggg	cctaaatgca	atccaccttc	ttcttatcag	aggaaggagg
157861	agcaagatga	gccaagggc	ggagaagggc	acgtacaggg	agagcagagg	tactggggag
157921	atgtcggccc	tgaagccagg	agcgaagagg	ccacaagcca	agcactgctg	gctgcccagc
157981	ggacctggaa	gagccccagg	cgcatectcc	ccagagggct	ccggagggctg	cgccggccctg
158041	ccgcacctctg	cgtttllgccc	tgttgaacct	gatttcagcc	tcagaaagcc	tgagaggata
158101	caattctaat	gtttcaaac	agcaagctctg	tagtaatttg	caggcagctc	cagaaaccaa
158161	tacaatgggg	aactctgata	tggactaaaa	gtggtagagag	aatatgagta	gaagaaatgg
158221	gctggggagg	gcagcaggtg	tgtcagtgag	ctaaacccct	ctcttccagg	gcaggaatgct

158281	aatggacaac	atcgaaaccc	gatacatcaa	gaacacgcaa	tagaatagca	tttcaggact
158341	tcattggcaac	cccgaaccca	gccagcagcc	taagagcctta	ctccctcagga	
158401	ctggcctctgt	aaagaggggac	gcacacgcgt	tcacactcat	gcacacacac	acacccgagac
158461	agaggggggac	ctggccagag	cagggggtgtc	agcagagggg	acgcacacgc	atctcacctac
158521	ctgcacacac	tcacacagag	acagagggag	acagagggag	agcaggggtg	ctacagagag
158581	ggacgcacac	gcattcacac	tcattgcacac	acacactcac	atcgacagag	agggagggct
158641	ggccagagca	gggggtgtcag	cagaggggac	gcacacgcac	tcacactcat	gcacacacac
158701	actcacacag	agacagagga	gggcctggcc	ggagcagggg	tgctcgcaga	ggggagcgac
158761	acgcattcac	actcgtgcac	acacacactc	acacagagac	atggagaggg	ctgtccggag
158821	acggggtgtc	agcagagggg	acgcacacgc	atccacactc	atgcacacgc	acactcacac
158881	cgagacagag	gagggccttgg	ccggagcagg	gggtgtcaga	gaggggacgc	acacgcaltc
158941	acactctgtgc	acacacacac	tcacacccag	acagaggagg	gctctggccg	agcaggggtg
159001	tcagcagag	ggacgcacac	gcattcacac	tcgtgcacac	acacactcac	agccagagag
159061	agggagggct	ggccggagca	gggggtgtcag	ctccagagac	aggaagtccc	agccagagac
159121	tgatgtgtgg	ggggggggtt	cotcacacac	ccacaccccc	tgctatagcc	cagacctcca
159181	gctctgggct	ctgcattccac	acggagggctt	cgactggggc	gcagctggg	agcacacaga
159241	ccggccagag	ccagggccag	gactgcattt	tcctgggctc	ctgtgcccat	ctgcgccctc
159301	ccacccgggtc	ggcctgtctc	ttgtcttgc	cagctcgcc	gaagcacctc	cctcgccgca
159361	ctctgggctt	ggcctggcct	cagcctctgt	gctgcacgt	ctcacttgga	gctggggcgc
159421	cgccagctct	agcttaagt	agcgccagag	agagtgcag	gagccgggag	cgccagctcc
159481	ctcccccacc	gtggcgtggc	gtggcgtggc	cggaatactc	agttttccct	gctttttgag
159541	acaaagatgc	taaatgcacc	gtcctgtact	cagagcccaa	gaagcacccc	ctccaccac
159601	agcgcctctg	ctcttgcagc	ccccccggcc	gtgtctctct	ctcacggccc	cacactctca
159661	ctggaacttc	tcctttttac	agatgaggtg	agacagcatg	cagagggcca	tgctgtcttc
159721	ccacagggct	acgggccaga	attacaagcc	ctgattctg	ggggcttcca	cgctcttcca
159781	ccaaagcctt	gtagccttcc	tgacccagg	tggggctcct	gtccccccgc	agctccccca
159841	accttacttc	cagcagacct	ccacctacag	acaggagctc	tgggcaca	ctttttttgt
159901	ccaggcccaa	ccagagctgt	caagggacat	aggaatgtcc	cttagctctg	ccccggagag
159961	tcaggcccaa	ggcctaccta	gaagttaact	tttttctgta	cattaaagca	ctccccctcc
160021	agagcctaat	cacagccagc	taacggccct	ttccccagtt	attaaagctt	ggaggaaccc
160081	tggaagtgtg	actgtatttc	tacacccacc	ccccccacaa	ttcccaaggg	ccccccccgc
160141	ctctctcttc	cccttgcccc	tgcatggagc	agggctcact	cactctctct	ctctctgttc
160201	tcctgggttc	ctgcttcaga	ccctccacc	taacacagcc	tcgtccccac	ccccaggggc
160261	tcctcccaag	acacgcgttc	acgtgcacct	ggacaagatt	aaataccctg	gggtggggag
160321	agcactgagc	ccacagtgtg	cttctggaaa	gcactgaggg	cacagtgttc	ttctggaag
160381	cactgagggc	acagtgtgct	tttggaagag	actgagggca	cagtcctttg	ggaaggtcag
160441	gtcgggtagc	cagggtgtac	tgcaactctg	ccagtgaagt	gggggtgttc	ccaaaggggg
160501	tttgtttccc	accattggga	gccaggtatc	caggttcagga	acagagcttc	gggtggggag
160561	gaagccatct	ctgggggactt	tcctcccatc	tcctctgtcc	cgcttcacag	gagaaagcag
160621	acttagaagc	tgcatgagc	ctctctagca	ccaccagcca	cagtcctcac	ctccggggac
160681	acgagctggg	atctgagtgg	cgcaaggta	atgtttctga	gagggccgtac	ccaggaccgc
160741	ccatggccag	cagaggaaaa	ctgaacaaac	gatccttcca	gagacaaagc	acagcccagc
160801	atggcccaag	gaactggccc	ccccaggagg	ctgtgtgagc	caaacgccac	tggaagaggg
160861	gcttggtggc	tgacccgaag	agcagaaacg	aggaactgaa	cagccacagc	accccgagag
160921	ctgggggact	tggaagagc	tctcgtgtgt	tgctaagaaa	ggagcagaag	acagagggca
160981	gggaagaaag	ctgcgagagc	tgggcgaagt	ccagaagaga	gtcggggagg	tgagagaggg
161041	aatatttcaa	aaaaaaaat	tacatgaaag	attccagcat	ggatggggtc	ctagaggcct
161101	tatacatata	tataaagaa	aaagctccat	agctagaccc	acgataattg	aacctaaagaa
161161	catcaaaag	aaagagaaag	ttcaaaagct	ccagagaggga	aaagacagat	cagatcagtg
161221	taaacagagg	ggcagcaggg	agacatccaa	gtccccagct	ggggcggtgt	gacccggcag
161281	cagaggggag	acaagagccc	aggaagtctc	gtggcgaag	aatctcaagg	ctcttgcctc
161341	gaggggaagag	gtgcccccat	tcaggagcca	gagccacccg	acagagagac	actgtgtgtc
161401	ctcagagagg	tcggcagagg	cagagggcac	accagagcag	cagagactga	gcacccggag
161461	aggggtgacac	ctlgggagat	gctgagagca	gcaggggcct	caggagcagc	gtgggggagt
161521	cgccacagag	tcaggggggtc	attgcatgtg	ggcatccagc	gctcctgctc	agggccctca
161581	gggtcccccac	gccccacaga	acacccctcc	cactccccag	gatggcattt	gagaccttct
161641	tgctgtgttc	ctccaggaac	ctgtgagctc	caagagaaca	gaagctgtcc	ttctcaaccc
161701	cagcttcagca	aggccagagg	cagagcaggg	acggagaact	gtgttcgaag	atgaagtggg
161761	gcagggagcg	agtgtgtgat	acctataatt	ccagcacttt	ggggagccaa	gaacccgtta
161821	agtcacagaag	ttcaagatca	gcctggggcaa	catagtgaag	ccccattctc	acaaaaaata
161881	aatgttaaaa	attagccagg	catggtgtgt	cacacctgtg	ctttcageta	ctttgagcca
161941	tgaggcagga	ggaatccatg	aaccagagg	tcaaggttgc	agggagctac	atgtgacagc
162001	ctgccttcta	gcccgggaga	gagcaagacc	ctgtctcaaa	aaaaggaaat	atatttaatt
162061	ttttaaaaag	aatacacagt	ggtgagaaag	gatggagaaa	agagagggaa	ggaaggaaga

162121	gataggggaaa	gaacaaaaga	aggaagaagaaa	gaaattccag	ctaattctca	gcaacagaca
162181	attacaggaa	cacaagagaca	aagccagag	cccaacccc	agctgggctg	ggtccccaca
162241	gcctccagacc	ctcctgtcac	tcgtccagct	ggcagcaggg	ccactctctg	gcatcctttt
162301	ctccagacac	ttggcgctct	tcctccagct	gggtttcttg	tagttctcca	taaatcatgt
162361	cgtaggtctg	gctcacagcta	tccttgtccc	tgaagatgca	ctcagctagc	agaagtcgct
162421	atattgttaac	acaatatcac	acaattccalc	gaagcctgct	cgagcgtgcg	cctcttgacc
162481	ccgallctctg	tcaccaacac	aacaggaaaag	gaycttgaag	accaggaggg	ccacagcaca
162541	gttccctctg	gagggagaga	gatgcgagca	gggtggcgga	gggtgcactc	atgagacttc
162601	tataagcagc	gaaccagaga	ccatccagcg	gagactgtgt	ggggactcct	gcacctcttc
162661	ctgtgccca	gaccatgggt	ccgcagcagc	actcgagtgt	tggtgttaac	tgcttcccc
162721	ggatggccag	cgggcacaga	gcataatgtgt	gctgaggagc	acatgcccca	cacaacccgc
162781	cccacatgtg	ctatgcatac	atacccccaca	cacacacgtg	tacatttcag	cacgcgtcta
162841	ggggaggtta	gcgtgccagc	accagctgag	actccctggc	cgactcatc	gccagttggg
162901	gagagtggag	cagacgtaca	ggatggaggg	tagaggtggg	gctggtccac	agtgaagggg
162961	ggaggcgaggt	gtagactggg	ccagggaag	gttctgaaa	cccagactga	aggttccccg
163021	gggtccctgt	ggctcatcca	ggcctggggg	ggagccagga	catgaggaac	gggcagtgat
163081	gaactctgaag	ccctgctggt	cagcagagcc	tcaccttcca	cgctccctccc	aaaggtccag
163141	cgagcttcaac	ctctggcacg	tcctcaacct	ggccatttagc	aaagagcccc	cccttggacc
163201	aggaggaacca	cagcaggcca	gccctcctgc	ccttccactc	caaggctcca	gcacctctca
163261	tggagactct	gagctgagac	actcccagca	gaaagggatca	tttatctcca	agcttaacta
163321	gccctctcca	attagttagg	ataaaaatgg	cagcatttac	tctgaagaca	tgcgattcga
163381	gagttattaat	taattgcgtc	gccgctgagc	ctcagcacac	agctgagctg	ccctctcaat
163441	aaaaagcatg	agactaaagc	aagacagggg	ggagaggcgg	ggccagctag	acaggggatg
163501	ggagagaaag	gagatgccgg	gcggggcggg	gacagccagg	gcgcccaca	gctggcctgc
163561	caggagggag	agggcccact	gcaaaagtgg	acagctgctt	ctagccctct	cacaagaagc
163621	cttgtgtggga	agaagcccaag	cacacgcaga	gggtgtcccc	gcacaccccc	cgcctccgtc
163681	tttttggaa	cttgacagtt	agaagacgac	aaggggccact	gtccagtggc	tggcatttgt
163741	ccccatgagc	ccacagagag	cagcagggaca	gctcctcttc	agaccacccc	aggtgggggc
163801	atgccaaagc	ccctcggctg	caggatggag	ctgagcctcc	ccaggatgac	tcctctgggg
163861	gaacacatgc	ttgtcgtggt	tgaagcgttg	gacgttgggg	gctggcctct	gagagcgccg
163921	gcccccgaag	gtcttccagg	aaggtcagac	cccagcccaa	aaagcagcag	cctgggcccac
163981	aaaagagact	ggtccctggg	ccagaagacc	ctcttccacc	atgcactctg	ggcaaggggc
164041	gtgtgtgcca	ttaggaacac	agctgcacgc	tatttctctc	ctgtcttctc	ccagctgaag
164101	atttcacaca	ccctctttat	gactgagggg	tcaccaacgc	tattccctca	acacacagcg
164161	ccccacacac	atgcattgtc	actcatgca	tcaccaatgc	agcatgcatg	cacccacata
164221	caaatgtcca	tacgatcac	atgcacacg	tgctcactca	cacgagggcg	tcacacagcg
164281	accatgcaga	cgcattgcaca	cacacatcac	tcataatgtc	atctcctgtg	acacatgcat
164341	cacatacaca	gttccacaca	tgtgcactac	acacaaatgc	acacatgcac	cagaaaaaatg
164401	ccccatgcat	gtgcattgga	gacattaccc	acacatacac	atgtgcactg	gtcatacatg
164461	tgatcatat	gcaggtatca	tatgtacagg	catgcagctg	acacatgcat	gcattcaagca
164521	catgcattat	acatgcacac	gcccgacatc	gtgcacacat	gcataacact	gatcactca
164581	catgcctctc	acatttgca	acatgcata	ctgacgtgct	cacacatacg	ccacacacat
164641	gcactcaact	acacccctct	gctgctccca	gectcactgc	ctgcttgctt	acttccccag
164701	ttctctccac	tacctgagaa	gggtgttatt	tgcccttttg	cttattgtat	gtcttaccac
164761	ccctcatctta	ccctcccgag	ccctcccatc	ccacactcaa	ccagtagaat	agaaagcctg
164821	ccctctctta	gaggcaaaaa	gctgtgaccc	tcctttccac	cagaaaacac	aggaagcatct
164881	cgtagcagct	ggaactcagc	ttaccagccc	cccgagctca	ttaacgctgg	ctccccagca
164941	cccaaacccca	ggtcccca	ggctcagcag	ttcacaccca	gctgcacaca	tcctcggcct
165001	ggcccacaga	atgctctggt	ccagactcct	aatgtaggaa	gctgcacctg	ccctgatcca
165061	agcgttgcct	ctctgcattc	aggccactcc	tgctgccacg	cactgagcgt	ccctgagcag
165121	ccggtgatca	gtcaaccctc	ccagcctctc	gtttgccttg	cagagccagg	cgactccccg
165181	gagagactgt	gagggctggtc	agcagcctca	ctgtccctgc	cctgcctggt	gaggggaagc
165241	agcccacgga	gaaattcccc	tcgcgtccca	tgctgcctc	gtgtcctctc	ctctcctgct
165301	gtgctccctg	gcagcagctc	cctccctcag	cctccaggtg	ctcccaatcc	tggcccacgc
165361	caactctctg	gaaacccctaa	gcctgcctgg	agtcacacct	cctccatttt	tttgggtttt
165421	ttttgagcag	gagctctgct	ctgtcactca	ggctggagtg	cagtgagcag	atttggtgct
165481	actctgcagct	ccactcctg	ggttccaccc	attttctctg	ctcagcctcc	tgagttagctg
165541	agactcagag	gttccaccac	tacgcctggc	taattttttt	tttttttttt	tgtattttta
165601	gtagagacga	ggtttcacca	ttttagccag	gatggtctca	atcctctgac	cttgtgatcc
165661	gtccacctca	gcctcccaaa	gtgctgggat	tacagcgcatg	agccaactgt	tgctggccca
165721	ctctctctct	ctgatagcac	agccctctgt	tcactcagga	acagactctc	atctctcggt
165781	ccacagtggt	caaggagggg	acatgaccca	gctagagcca	catgaataac	gttgtcatct
165841	gaactacatga	acatagctca	gggatgagga	cttgccacaa	gggtattcat	gaagttaaat
165901	ccaggaaacta	ttgctggaat	tgttggaag	gaaaagctgt	ttccccccgg	gttttgcctt

165961 gtgaggagggt aatgcagcag ctgcctgcac ccatTTTTct gcctcagaga agagcaggcc
166021 tgagagtgaaa atgcacatag aggaagacaa agccaaaggtt tggggatgtg ggaiaaaaaaa
166081 caactgaaaa tttatttggg cactctggatc cagccacacc tgaagctaga tagtctcaga
166141 tatTTTTctt tttgcagatt aaccaataag tTTTTtttct tTTTTttctga agccacttll
166201 acttggattt tcaacacttg aggaataaca cctgactaat cttctctctc tctctatagt
166261 tcaaacctctc catctccccg atctctcatcc tgcgtccctg agcaaacctca agccacactaa
166321 aatctctctg gctgcaggca gggccacaca cacatctaca caagccgaca tcttaaacct
166381 gcaatcacal ctaagactgc acacccctcg cagctgaltg ctgccccacc acagcatctg
166441 tgcctagcgc accagcttcc ctcaactccc ttgctctca ctgccccctc actgggtccg
166501 tctagcttcc accaacctccc cagcgacctt gcaaaagccc ctcttggtg ctgggctcga
166561 tgcctcctgc catcccactt cccctgaacc cctctcttg aacctctctc atccttttagc
166621 aggtggggag cccctgctct aagaccctcc tgcctctctg acttggtctg acatgtgtgg
166681 ggggcaaggga gcacttctct cccctctctg cctctctctc ctgttggtgc cggccaccca
166741 cccgcaaggga tgggaatcac agccctctat acagagctgc cccaaccgga atctctacca
166801 gaccttgcct ctgcactcag cagctgcatt tccagctgct taccaacctc caagagacag
166861 cccagctgac aaacactctc tccctcagct ctccctcagc cccagggccc acagaccxaa
166921 caaaagctgg ggaagtgttt tccctcagct ctccagctct cctgagtaga ttatgtgtgc
166981 gtgctcgag atcccgccct ctccagtgtt cctgagtaga attatgtgtg ccatgtgtgc
167041 gatgttccgg cagaacactg gatgatccag aaaaatttcc attatgtgtg ccatgtgtgc
167101 aggcagatca caaagggag gtacaatacg accttttgt tttatgtgtg gtttacactg
167161 actgtcatag acgtgggaga cacacaaaaa tggtaaagc taatggactt gcaaggattt
167221 ccttttctct cttttgctt atttctgcac tctgttttat tctgttttat atcagcagt
167281 ttttaactta aagacaggac aaaaagtctg cagccatcgt agtccacccc aggaactcgc
167341 gtgctggaaa tgccaaaaatc tggctgggacc gggagagga cagagaggtg
167401 caccaccgag ccgtgctctc tttgaatcac tccctcaggg gccactctaa ggtgagact
167461 ctgaagctg agagaagtga aataaccgag ccatggtcac agcaggggta gctgaggttc
167521 cagtctctcc tccccaccc aggtcccgac tacagctctc gatccacccc cctcctctg
167581 gacaaacccc gggccacgct gggctctccc gcaagccctg aataagactt ccatggccac
167641 atcagcaccc gggagagcgc agggagccgt gaggcaatc cagatagact cccaaggtc
167701 cggccaccac agagtcaagg aagcctctgg ctgcctctg aataaagaa gtccagctga
167761 agggccctct gggagggcac ctccgaggg cctcgccagc tccgcagcca ggtggggccc
167821 agggccggc ggcagaaaaa ccaaagtggg cagtcaaac tggacagaaa agggcctctc
167881 tggctctgag ctccctcact cggggcctga cggtaaccag gggagttaa ctctcacaca
167941 caagctggcg cagaccctct aactgcaggt gtccctaagg ggcagtgcct gccaagtctc
168001 ctctctggag tggagctctt cctcctgagc acggtgtgga agagggccgc tgcccagcat
168061 ggggtgcgtga ggaactggat ctgcctgcag acaggactg cgggagggca gctgcgaagg
168121 gcttaggact gaggccactc cagagccaag actgacagtc ccaacaaac caagggaccc
168181 tcactgtgct catgctggtc aggggctgca caggaagtga cagagggctc gtgcctcga
168241 gaggcgagga ccgagccagc cagtgtcaag gacgaggtct gcaactctga tggaaaatcc
168301 atccgtggtc cccctctga caggcccttg gctgccttcc ctcttaataa gctgcgaagg
168361 gaaatggcct ggcagaaagg aagacaatgg gggccacagg gcagcagtcac aggagggcct
168421 gggcccccac cccctcagct gtaggactcc caagctaaag acagtgtgaa tacttgggt
168481 aggagaaagg agcagacacc aggcctctgt gcaggtgtca gcatggtgta gggcagggag
168541 aacagccgac agcttgatct ggacacctgg tctctcaca gggccacgca ctctgccagc
168601 gcggttgaga cctgtcactg cctgtgggga gctccttgga gagtcttaa cctgggctc
168661 ataaactggg aaaaagaaac acttcaactt catcttact actctgggaa ctaagactgc
168721 catctccttc aactctgag agagcgacga accccagact taacagcacc tggccacctg
168781 agggcagaaa tcacaggctt cggaactgt cagcaccgct gctgcagacc agaccaatct
168841 gtgtttacac tcagcacaa ttaaatata gaggtgttaa gatgcactcc taggtcttat
168901 taittaata atcaattaaa agcacacaaa aaagttttac atgaatttct cctggcagat
168961 taitcaaaag actaaaagg tggaaaacac caatccatc aaccatagag tggatttaaa
169021 aaactgtggt agctccatgc aatggaatca tatctttcca tccaaagtga tgaagtctgt
169081 atcagctgtc caacttagat gaacctgaa aacatlatgc tgaagtgaag aagccagaca
169141 caaaggcgaa atattgtgt attcatltat actaatgaaa tgcccataag agcaaatct
169201 atacaggcag aaggtaggtc agtggttgcc taggcctggg aggagcgggt aattgggggt
169261 tactaaaagg tctggggttt ctttcaag atcttcaaat tcttgaacat gatctaaagt
169321 gtgacacat tctgtgaaca tactaaagcc attcatgaa ttttactgta taagtgtata
169381 aaatcacat tatgtgaatt atgttccaat aacatatta acatatttt cttcttttt
169441 ataatgata aaaaatttaa aagcatgtgt cactctatag ccccaactgg acttactctc
169501 ttttttttga gacagagtct cactctatag ccccaactgg gctgcctcag ctctgaata
169561 gctcactgca acctccactc caagggttca aacaattctc cggctagctc atctgaata
169621 gtgggaacta cagatgccc ctaccacgccc cggctagctc gctcaggcca tctcctgcca
169681 ggggttccacc atgttgcaca ggttggtctc aaactcctga gctcaggcca tctcctgcca
169741 ttggcctccc aagatgctgt gattatagc gtgaggcacc atgctggccc gatagcttta

169801	tttcaatcag	atttattttc	tttgttattt	gtagtgtaca	ctaaacacat	tattttgtac
169861	aggggtccat	gggctcacc	agggctggac	gaggggtcac	gacatggaa	agctgtgcta
169921	actaccagga	cgaagagac	gaaggtttgc	tggagagagg	gtatgagaaa	gttggccagg
169981	ctgtctgggt	atgcagattc	caggcaaaag	ccagaggggt	ctcaggagacc	ccccactctc
170041	tcacacccgt	acacacacta	ggccttggct	ccttcaagtc	caggtctctc	agccaccocga
170101	aactctcagc	aagacaaatg	gtccctctct	gcaacctgca	cgagactcgt	tccaagatcg
170161	ggcagagtgt	acagttgatg	gggtgtgtgt	atctctcgga	gctctcgtct	caggccocaca
170281	ggatgaacgt	gaaggtggag	atgggtggga	gagcaagcag	gaatatgatg	tggtcaggag
170341	gggctctgaa	tgctgggtct	gggggctgtg	gcttlaacct	caaatgtggag	gagccatggg
170401	ggagcatttg	gggtggggac	ggagagggga	ccgcgtgctt	atcatggggca	gctgggaagg
170461	agtgggagaa	acggggggcag	ggagagagaca	gggcagatgt	cttggagagca	gaaggcaalg
170521	gggacccata	gtctctctct	acacatcgac	gtggcttttc	cttgggggcc	atcctctccg
170581	gcaccaactc	ccagagggtc	agaggacaga	ggcagccctc	ggagccacag	tggtctctgt
170641	acagagcctt	ggcagccgaa	gtctctctcc	gctgtctctg	ccatagtgat	ttgctgtgaag
170701	catgtgacct	aaactgaacc	agtgagaact	caccaactgg	ctctcccgag	aaacctctgg
170761	gaactctggc	ctctctctgc	cgggcgctgt	gtgcactggc	caggcaggag	ccacggggag
170821	ggcgctgtag	gtgaaacctg	gcgggagccc	tgggctgcgc	ccctgagccc	cgcatctccc
170881	ctttggcccc	agccgtctct	cccccggttt	tttgtcaact	gcagtccacg	gagtgtggac
170941	ctggggatgg	ccgtggggag	ggacaggggc	acaggcacca	cggtgacagg	ttgtgtgtgg
171001	agggacaggag	cgaggccagc	cgctctctgg	ccacaccttc	cggaatctact	cccaaaaggag
171061	aaacggccagg	cgcccaagct	ctcccaagct	aggggtgcgg	aggaaggagat	cagagcgctc
171121	ccccaaacaa	ggctgcctct	cgctctctga	catccccctg	ccacactgac	cttcatttcc
171181	cttcttgggg	ctctgcccgc	acgagactct	ttccccatct	ccatgtgaa	atttttctgc
171241	acaccttact	aattgttttc	aggtggggaag	gattttgtgtg	agaaattgctg	tttgtgtgtg
171301	cggtgtgtgtg	ttgtgtgtgtc	ttgtgtgtgtc	cacccccagc	cgggcagctg	ccccaggagt
171361	gctgtattgtg	cgggaaagagc	tagaggaccc	ttggcctcag	ccatgtctctg	aggggctctc
171421	ggtagatgta	cgggctctctg	gcttcttcca	gtaaacatct	gtttcttattc	tgaccacaga
171481	attctatttc	agaacagtaa	cgggctccac	tggaacacaa	ctccagggccg	gtgatgctct
171541	ggggcaggag	agcatctctac	ccgggctgtt	gcccgcgcgc	cgcccccgc	ccccactcgc
171601	atggcccccc	aagccccgcc	caacagagctc	ctaaaaggga	caaaattctgc	acactggcttc
171661	gtgtgtccaa	cgcttttccac	actgtgtgttc	ctcttagcag	cgccatcgag	aaatattgat
171721	cttgttagtaa	caggcccttc	gctgtctcag	gctggccctc	gcttctattg	atctctatca
171781	taaccacaca	gagtagttac	gcctatcagc	ttttaccttt	ccacagatga	aaacaggagg
171841	ctcgtctccc	caggggaggga	gggagcagaa	aaaccaagccc	acgttttgga	gcgagatcgc
171901	gtctgactcc	agcgcacggg	accttgacag	ggctgatggc	caaaaattgta	agtcattggg
171961	aggagaggga	gaggggagag	gagagagagg	ggaggagaag	atgacccgttc	caaaagccat
172021	gcagcactca	ttccaaaata	agagaaaactc	tcaggtgcct	ccagggagag	tgaaatgcgc
172081	tgaagccccc	cagggttaatg	ataggggagca	gcagagactc	ccaagggcacc	tggaactctgt
172141	ggacccttct	agaggtttcc	atccttgctc	agctctagcc	ctccaggcaca	tagggccggc
172201	agattttctg	gttctcaaga	gaaactggaa	atcttgagct	tttgtgagaaa	ttttcaaaac
172261	ctggaaaaag	caacaaaaac	tcactgtgtg	gtggacatgg	cccgtagcccc	accagttggc
172321	aaactccggc	ctgcacaaatg	ggaaacctga	agtgacagaa	attatctcat	tgtaagggtgt
172381	aaggcaggga	ctcaaaaacg	ggctctctcag	tcacatctca	tagactctcc	aggtttgtacc
172441	accgaaatct	ctgcgcagcc	ttctctctga	atgtgtctac	tgaaccacat	cccccgtagc
172501	aactcagggc	atctttcaaa	gaccgcagtt	ttcctggcag	ccctgcacaa	acctctgtga
172561	gtcccggtgc	actgctcccc	aaccacaggg	gacacctaac	agaccacagag	ggactgagag
172621	agccctcttg	actgaaaggg	aaggtcaacc	gcgcacactc	atccccatga	aaccacacaa
172681	tgggcacatc	ctgcgttctga	agctctactt	gatgcagaag	agcaggggaaa	ggaaagccgt
172741	gcaccgctgt	ggaaatgggc	ctgggagatc	catccaggcc	aagggttggc	cactttgggt
172801	ctaccaccac	cagcattctac	agaagaagca	gcacctggtt	ctcaactccc	tgaagaaggc
172861	ggctagatgt	ttgaaaggca	tggttgtagc	ttccacacaa	aatttactaaa	gacataataac
172921	gaaacatctg	tgaaaaagaa	gagacaatat	tatgaactgc	agaaagttaac	tgccacagct
172981	gagaactctc	acccccaaaa	acatccaaaa	aaaggaacca	aatataataa	aggaagtcca
173041	taaaaagtct	agatgcaaaa	taaatagaga	aactcatata	ctctgggaagc	gaacatcgct
173101	aaacatccaa	cgagagagatt	taggttttgt	ttgttttgtt	tttttggaagc	caactctcgt
173161	ctgtgcacca	ggctggagtg	cagtggtgtca	agcgtactga	gctcactgca	ggctccgcct
173221	ccccgggtga	cgccattctc	ctgcctcagc	ctccggagta	gctggggact	cagggccgca
173281	ccaccacgcc	cggttaattt	tttgtatttt	tagtagagag	gggggtttcac	cggtgttagcc
173341	aggaatgtgt	cgatttctgt	acctcgtgat	ccaccgcctc	gggctctccc	aagtgttgat
173401	atctcaggcg	tgagccacag	caccgggcaa	gagattttat	tcagaaaggag	tttttgtgat
173461	tatctgtgtg	tcataattgt	gccattctag	taacagagag	actcgtttgt	tgaagactat
173521	gaagatgcag	gtcatgtgga	gcaaaaggatc	tgaagcaggg	gctgttgcca	ggaaaggcgc
173581	gccttgagga	cagcaatgct	ggggcttggg	ctgtccaggt	caaccatggg	tccaagtgtg

173641 ttctgtggcag ctctgcacat gcatgtatga ttgtgactaa tgtgtctatt ttaactctgt
173701 ttctctctgt ggaataaaga tgctaaatct ctgagaatat catltagctgt aaagaacgag
173761 gtccacggta ttctgagtaa taatcaacct actgtattat aaatataaag tgttccacgt
173821 taacaattct gtgttaaaata taatggaaata gacctattgt aaatgtggaa tgttccacgt
173881 agtctcttag taataaacct actgggtact gtttttaaac acttgaaata taactctctat
173941 ttccgtgaga aaattaaaca tcatggaagt tctgttcaat tgttccagaa agatttctgt
174001 agggcctctt gcaaaagggga cactactatta ggcatttttt taaaacctca cgtatgtgat
174061 acttgccctc tcagatgtca gaacataatt tgaaccaaca gcaattaaag accottggac
174121 tgggtgaaaa gaaaaacaga aaaggccagg cacagtggct catgctctga atcccgacg
174181 tttgtgagac caaggagggc ggtacacag gtccagagat ggagacgag ctggctaaac
174241 cgttgggaac ccgtctctac aaataataca aaaaatttag cgggtgtggg ctgggycgoc
174301 tgtagtccca ctactcttg agcgtgaag aggagaatgc ctggaacca ggagcagag
174361 ctctcagtga gccagatag tggcactgag ctccagcctg ggcacacag tgagactcgg
174421 tctcaaaaaa aaaaaaaca acaaaacaaa acccagaaga gccagaata atccccaat
174481 cctccatgaa ctccacagat gatggaatat acattccaaa gtcagaaga aaagatgaat
174541 tagcaaacga atggtgttga gacaacagca tggcaattta gcaactttaa acactttaa
174601 accaaacact tatacagat taagaacaca tttaacttga attaaaggag ttacattat
174661 aaaaaacag tcaggaaaaa agcaaaaacat ataattgagt atttcttga aaagctgata
174721 cttttctaca ggctgagtaa ccagaagaaa gcaactagag aggctgacag agctgaccga
174781 agaaaacaaa cccagaccgg gcagaataaa aaaggaaact ctgcaaatat aaacccacca
174841 aagctagcag agggttaat actagcccca gggagaacca atcagctca gaggaggaaa
174901 gtggggggc ttgaaccaca gccattaaag aaacaagttt caggccgggc ttgtgtgtct
174961 agcctgttaa tcccagact ttgggaggcc gaggcgggt gatcatgag tccagagat
175021 gtagccata ttgctaaaca ggtgaaaccc cgtctctact aaaaatacaa aaattatgcc
175081 gggcgcgggt gcgggcgcct gtatgtccag ctactcggga ggctgacga ggagatgtgc
175141 gtgaacccgg gaagcggagc ttgcagttag ccgagattgc gccactgca tcccgactgc
175201 ggctctggcg acagagcgag actccgtctc aaaaaaaaaa aaaaaaaaaa aggaacacag
175261 tttaacaat aacacgattg cctgactccc accaagggtc ccaggccacg acccccacg
175321 cggactcctg caggccttgg gaggccaggt gggcactctc aggcaggttg ggggaaggct
175381 ttgaggaact tgagcccccga gacttctaca ggacagcgt acacaagaac aggaacggcc
175441 tacatctaac caccctatcc catgcccagt cctggaccag ccatccctgt ccccaaacgc
175501 cacagctgac tcaccacagc tgagactgccc ccaatcaggg ccaactgaa tgttccagc
175561 cctccctgca actgcccggg ctttggtttt taatgtccct agtcccgaga atgcagaacg
175621 tctgcaaaag gaagaacgca gcaactctgg caacctgact gttggaggag accactggca
175681 cacatttccc ctccccagc gccctgtatc tccaccagc gctgtgggacc ctccaacctc
175741 cactccattc actcccagc ggagcaaaag ctggcctcat ttggccactt ccccagaca
175801 ccactctgac aaacttccca gcagccctgc tcatcgggccc accctcccca gtgagacac
175861 agcccgtgac agcctgggca cccaacacag aaagacagcc caaaggccca gccgccctat
175921 ctgagtgaga ggaatgctct gcaggcgagg gactcacgct tgaccttggt tgttagaaaa
175981 tgaccacaaga actcagtagc aggccaaagg gagagacat gaahtccct ttccagcaat
176041 tcaacctgct tgtacagtg gtgagttatc tgtacctctg agggagtgaca gatttcatct
176101 ggtgtctcag ctgcagttaa aggagctggc tctggctctc tctggctctc gtttgggta
176161 tcccctcagc tgtccctga ttctctccca gcacattggg atggctcgtt cctgttccc
176221 agaatgtacc aggaaggtac aggtgcagag tgcagctgac gacgaaggcc cccagcagc
176281 gaagggtccc caggagtcct ggctgatgaa agactctggg aaagaaagggt gggaacaggg
176341 agagcacaaga agggcgaggg aggggagagg gaggagaggg ggaagtgagg ggaaggggaa
176401 aaaggggaga tgaggaggaa gagaagggaag gggagtggga aaaaaggagg tgagaaagaa
176461 agggggaga gaagaggagg aaggggagga gaagctggag ggaagtgagg ggaaggggag
176521 ggaagaggag ggaagaggag ggaagaggag ggaagaggag ggaagaggag ggaaggggag
176581 ggaagaggag ggaagaggag ggaagaggag ggaagaggag ggaagaggag ggaaggggag
176641 ggaagaggag ggaagaggag ggaagaggag ggaagaggag ggaagaggag ggaaggggag
176701 ggaagaggag ggaagaggag ggaagaggag ggaagaggag ggaagaggag ggaaggggag
176761 gaccacagca aaccgtctca tgacagtgtg cctctcagct ccatattagg ctgttctgga
176821 aactcaaggcc ttctccctgt cgcgagggct gagccttgtt ctgggaacct ccttgttgt
176881 atcagagccc ccaggctggg ttgagttccac ttgagctctg tttaactcac aaatgtcttg
176941 accacaacca tgtggtttcc ctacaccaa ccagtctccc accctccaga cgttagtgtg
177001 gtgcccacaa gttgagtga atgtcacgc caactactg gggctcagtt agccctcagg
177061 ttaagggcct actccacac agcagcccca ctgcccgtg cgttgacctt atttccact
177121 tgcctctgta ccaggcggtc ataaattggg cgggtcccaa aactccctct ctaggtttgt
177181 taatttctta ttagtttggt gcaaaaagtaa cttgacctt ggaactcagc agacatgtgt
177241 tcattaatag aacagatcac agaactcagg gaggcatatt acttactatt tccagtgtg
177301 tattaagac acaaccagag aacagctgca tgggaaggat gggcagggca agaatcag
177361 aagggtcga gagcctccac gccctctctg ggcatccacc ctccagagaa ctccactatg
177421 tcaccaaccc aaagctctcc aaaccttgct atttaggggt ttccatggaa tctccatgac

```

177481 tcccatcccc agaggttggg gaagtgaagg tgaattctca acccaacaat ccagaaactc
177541 ttatatcaag acaaaagggg ctacagaccaa atctgataat tctatcaccc
177601 ttaaacctca ggaatttaca aaggttttag gagctctgag cgaggaaggt tctatgaccc
177661 acaaacatac atttataatt tcccatatc acagggtctc gcaagctctg gactctctg
177721 acccaacctg cattgcccca gaagcccgaga cctcagctca gggccagagt gctgggatca
177781 gtggccagtg gccagccatc gtgcagagtg gggagggcag agggagggg gggaggggat
177841 gggagagaaa gggagggagg agtgcagaga ggaagggagg gaaaggaagg gaaaggaagg
177901 agcaggagag cctccagagg ctctggcacc tgcctggagg ggcactgagg agccacaagg
177961 aggatgctta ccaactgcccc caocttgccc tgactcaggtc tgcagagtca gtgctcaacg
178021 cccatggagc tcagggtggc gcaccaagtg gaacagctcc cgactcaagg ccagctactg
178081 cccctgggagc cctcaccccc ttatcctgcc cgcttgctgt tgaggacccc tccctactgt
178141 gtggccttgg ggaatgccat gcgacggcgg ctgctctct tctgcgggg ccttgaattt
178201 ccagggtgaga gacccgagag cagaggggcca ggagccgaca ccgctgtggg agcttccccg
178261 atcccaagcg gggccatgcc cactggccat cgcttctct ccactcaagc cccccattg
178321 attccctttg actgaggcca gccagctcca tccccctctg cctccctact gacacacac
178381 ttgtgagcga ggtggtgtct taactgggaa aatggagggt cacagggcaq caagagaaga
178441 gcagagcccc aagtctgttc tgtgagcccc cactccccac agccctcgct agtgagcccc
178501 cgtctcttgg tccaaaaagc tctcctgtgt agcttttgaa aagtacagac cccagggttc
178561 accgactcgt gataaagcag aatctctcag gatggggcct atgtacagat gctgtgcmaa
178621 gctctctcaa tcccatgcc atcagtgggc aggcagagag atacacaaat ggaagtgtg
178681 ttctgcagaa caagccacag atttggtttt ttctgttttt tttttttttt ttttttttta
178741 cacagagctc caactctgct ctgtgctgga gtgcagtggc gcaatttcag ctactgaatt
178801 gctccgctct ctgggttcaa gagattctcc tgcctcagcc tcccaagatg ctgggactac
178861 agggcttggc caccatgcc agctaatttt tgtattttta agtaagtgtg gtttccaca
178921 tattggccag gctggttctca aacactgcac ctcaagtgtat ctgcocgctc cagctctctc
178981 aagtttccaa ggcgtgaagcc actgtgcccc gectacagat ttcttctaa gagggaatg
179041 tctttgcttg gaaaaaaccc ttctagaagc acgtgggata acagataaga ctggggactg
179101 aagagcagga gactctgcag acagtgccac cactcagag cacaggggcaa agtccgtgt
179161 taagacagga ggggagaatg gagggattgc tccagagaaa ctggacagag gggaggagag
179221 gccactcgag gggagaagaa aacccacaga tgaagtggaa tgcactccca cacatggccc
179281 aaagggcttt ccaagaacag ggtcctgttc ttccagaagt caccocgctg ccggggatg
179341 gctggctccc ccacttctg caggaaaggt gccctccctt agcagggtct cccatgtgtc
179401 tgaatgggg actgggtttc attcctggcc cagcaaaagt gacagctgag gccgaggtg
179461 agaacacaaa caccgccgaa tgcacaaccc tctgctttt ttactgtgaa gaggggagc
179521 ttaagagtta agacacattt tatcttgatt tcagctcctt tactgataaa aatgggatct
179581 tagaaaaaga gaattcagag gatattgctt tcacatgaaa tccactttaa tccagaccaa
179641 gcgctgtatt aagcagagaa ataccggggg cgttgcgctt aaagccagcg acagatgagg
179701 atgcgtgtag ccacggctgt gacttaacat cgtctggaag gttccagacg gtgcagattg
179761 gagatgagga tccaaaatta tcagtgttta cagagtgcac ctgtatacct agaaaaacca
179821 agggactcca caaagatata ttagaattaa tgagtccaat tgactgcttt gataaaagt
179881 agttctataa aaatcagtag ccttctctgt taccataaga tggaaaagta tcaatgagca
179941 atggcaaaaa gaaaaaaaat ccttgaataa atcctgattt ttaaaaaagg atgacctgt
180001 aaaaaatgt tatataataa acttcatggt acaacatcaa acagaaatga aataagctga
180061 taaaaatac tatgttctgt atgaatactg taaattatgt caactctccc ttgttaagag
180121 ttctaaagtt aatttactat caatcaaaat cagcagggat ttttctaccc gacaaaaaca
180181 ttaaaaaatt catttgaaa gataaataca tgtgaccagc caagaaaaat ttgttaagag
180241 ataccagcaa agggggagct atccagccag atgttccaaa gccacacagt ccaacagaaa
180301 tagaactcga gctttatatg tagccatggt tttaaaaaaa aaagacacag tgcgaattac
180361 ttccagaatt atattttatt taacccagta tgcctcaaat attaccattt caaactgtaa
180421 tcaatacttt aaaaatgatc attaaacatt ccaactctt ttgggggggt agcaagtctc
180481 tgaatccaa tgatatattt atactcaggg taacctctaa ttcagaccag acacatttta
180541 agtcttcaat agccttgtgt ggcttatagc tactgtattg gacagcaaa gttctcaagt
180601 cttaatactt tgactattat agccttacag ttctataaat aaacacagct taattttgt
180661 acagaacacg aaagatggaa aagttagaat tatctgcggt actcgttaac ctgagttgt
180721 atttgcaact gactggcgat ggtgtactat aaccataaaa atagacagat ttaattttgt
180781 accaatatgc tcacagctct caagccccct ttctcatatg gtccagaagc tctgtcccaa
180841 tcagaatgag cactagcaag ccttagacaa gccttgcagt aaactgtagg acaccctact
180901 ttccatgtgc tcccaatccc tgccttcaaa cagcgcgctg gggagggtgg ccagggtcag
180961 gtaacgggaa gggctgcaga gcctccacgc cctctctggg cactcacccc ccaagacact
181021 cactgtctca ccaacccaaa gctctccaaa cctgtctggt taggggtttt ccaggaatca
181081 atcccgggt gttaagatga caggctttaa cacttggaaa tacttttaat acttgaat
181141 ttcagattag gcttcaaa ggaataatc acaggtttaa agatgaatc ccgactctc
181201 gcacagagct ggtgagggtc ggtccctcca acacgcatct ttccatttcc caacccctga
181261 actgaggtg caatcgcgcg gtctcaggcc ccggccctcg ccccccctcc ctggggccag

```

181321 acacttccaa cgcagtcctt cactgtcaac accctggaca ctcaaaatgt ctttgaagag
 181381 tccagctgtt aggtcaaaaa gttagggtgc accgacaaga gcgcactgtag ctttgcattg
 181441 gtgaggcaga taacctatccc accgtactt cgttaattcaac aagaggaaga
 181501 cccccggagg cggagcgaagc caaccgaact caggttccggg gagccgtgac aagcgacggc
 181561 caggcagcag aatggcccca cccattaggc tgcagacagc atcagagaaa ccaagcagga
 181621 tggcagtgagg gcccgcgcct ccagggaacc acaaggcaca gagttgtcagt ttaaatggga
 181681 gtgtctctcc ttctctgggt cgtataaag cgtataaag tctctgaacc cgtggcaact
 181741 tagagctaag gatattgcaga cagcgccacc gcttttccac ccccgggcct ggggtlaggc
 181801 ctgtctgcaac acccaccctt tcaactggac tgaagtgcgt gcgggacgaa cctctctttt
 181861 gactattgag gaactaaagtt tctcaagalc gtgtgtctgt gcggccctgc cgtgcacagc
 181921 gcgcacagcg gacatattgc ccgcgaagac cctccagccc acaataggcc cctgctctcc
 181981 ctgggttaac cgtgaaatg cacaaggcac ggtggggatg gcgtagagcc tggccaaagt
 182041 cagcagcacc caggccttgg tctctgacac taggggtccc tctgtcctct cagggaaagt
 182101 gtgcacccca ctctgcagcc ctggagtcgt cgtgcagtg aattgaaatt catggcaag
 182161 cccgaggagc cagctccggg ctctgttaata acttctccat attaatggca gccccgcaga
 182221 cctgtggagc agctgatat tcaattgaaa catatcgac agagtgtgca gctctactct
 182281 ggactctccc attccggggt aaccacgaat aggggagtgg ggtgtctaac gtggcgccaa
 182341 gattgaaggg gctggggctc tctctcccc tgaattacct aaccocacct aagcgccctc
 182401 ttctacagag tccaggaaac ccttcccagg aatatataa acacattgctg gccactgctc
 182461 actcaaatgc acaccccttc atgcagacac actcaatgtgc acacacgtgt ccaacactcat
 182521 tcaagcgtgc acaaacatat gcacccccac agcgcactgt catgcacgta ctctccatgt
 182581 caagacatgc agaggcaaca cgtacatgca cacactcccc cgtcagcaga gcccagagaca
 182641 cgcacacgtc cacatgcagc caccagcctg ctgtgcggcg agcaacacag gqaccccagc
 182701 gctgtgccaa cccagagagg caggtggggg aaagaaggac atgtctgctt gcacagaca
 182761 acccagggtc ccaggagggc tgcattgcaac acttgcctct ttgggtctctc agtatccaca
 182821 tggggcaggc ccagaccata gctcaggcac ccaagcgcca aagccagcct cgtctgtgag
 182881 cggagtctcc ttgctgattc ccttactcac ccagggtgtg cccctgtgag cctcacaaga
 182941 ggcgctgggt aggcactggc ctcaactctc cagctgactc tgcagagggg cctgaatcca
 183001 ctcaactccc cagggtgagc cctctgtagc ctccacagca gccctgggtga ggactctggc
 183061 tctcaactccc agctgactct ctccaggggc ctgcaattcac ctgcactgta ctgccttagg
 183121 actggagaca tgggggacag ggaagccagc gcagagccatt gggcgagaga agcctgccac
 183181 acctgaatga gcaatggccc gcgcacattt accagtgcga gaaacatacc agccacacac
 183241 atggagtgca gctctcccag ggagagctgg ccttgggtgg cctccggctga gaccagtgtc
 183301 gactatggcc gccttcaatg ggggctggga gccggctcca cctctgccct cctaagaag
 183361 tctcaggaga gacgccaagc cacaacagaga acctgcataa gtaactctcta gggcctcggt
 183421 ttctcacttg taaaatgggg ccacggtggc tggctcacag tggctcacag gaaaaccgca
 183481 taaaagccccc aagcagaggt agatggtgtg gatgctgctc ccccagcccc ctctctaggc
 183541 tggcccgccc ctccagaggg actgcgctgc actgcctggat cctcaggca tctctgtagg
 183601 ggcctctcag tcaagtgaag cagcagggga gggagtcaat gccccacacct cccagggggc
 183661 cctgagggga acgctggagt tctgctctct gcagccccct ccccaggagt ctcaatccag
 183721 accagcagag caggcgactc aggcggggagc tggaggggccc tgaccagatc acgggggctc
 183781 tgaagtcacg cagcagtggt ttgagtgttt ttaagagcag taagaacacg gctagcaggt
 183841 taagtgaagg cgtgacatga gttgctgcat ggtttctaaa aatccctcag cgtcgagagt
 183901 ggaagaagggt gtgcggggag agacatggg gcaggtgggg gtaacaagcg gtacaagcag
 183961 ggtgtggccc aggcctggag ggaagccggg cctgagagga gctctagaat tggagctgac
 184021 ttggaggagg gcatcctgga aggggtgggt ctggaagaga tggggagag caccgtgact
 184081 tgagtattgg gtcatccagc gaaggggggt ccagagagca ccagcagggc aagcctggag
 184141 cccagagaga ctctcaagcc caacatgggg tctggagtcc agcagacagg tggccacaag
 184201 cgtgttccac accctgcaca attctctccc ctcttgtat gataggcccc tctctattcc
 184261 agagtgtggc agtctgtccc tccccctcag tcccaagggt gggcagagga ccttgctagt
 184321 gccgcacca cagccatggg tttgaagatg ggccccaatt cccctcccca acctcacaag
 184381 gcttccacca gaggaggagg ctttccatct caaagggttg cctctggaaa cctctgttgg
 184441 gcctggagcc gctggtgggc cgggggttga ggggtggaga gcaattctga atgaacact
 184501 gaggagcagg cgtgagaaat ggaatgacat gctctgagccc ctctgagccc caagctgta
 184561 agccaacact cctcaggacc ccagataat tggggccatg ctgtcccttt ttgtttcat
 184621 ccaatatagc cagggggtttc tgtcacttaa actgacaaag tcttaaggtg caaagtatg
 184681 cggagctgta gactggcctt agcagctgtg taggctgctt gcagctggg cccagagga
 184741 ctctgctccc taccctggga gcccccagac cctagaggtc ccaagaggtc ctgcagccc
 184801 tctttagctt tcttgacaac ctctcattgc gatccaggca gtgcccagggt ggggtggag
 184861 gaagggaagc cctaccagcg gcagctgcga atctcaggtg cccctacccc actgcacaac
 184921 tccgagggcc cgtcaccacc tgcctgatct gctctagcac ctgagcctgc tggagccag
 184981 tgtgtggcct cgaaggcagc agccttgcgt ggaaaaatga aactcatctc gatccctatg
 185041 gccgctctct gacttctctc tgtggccttt atgactctta atctccagca agacagaggt
 185101 ggtgctctgc ccagacagca gccccgacgg ggaagccagg cctgaaaggc agcagccctg

185161 gattttttct tgccatgcaa ccttgggtct ctctgagcct cagccacgca gtggggcagc
 185221 cctatctcac aggggttggt gtttcgcatt aagcatccag ggaagggcat acacttccct
 185281 ctctggggcc caggattgga ccaggccctg gaggggcgag calctccacc gccacgccca
 185341 tggaggcttc atccactgct gccagcctgt cgaggagagg atcccccgtgc tgcctgtgag
 185401 ggggaggagc agggcgggca ggttgcttcc atcccacgga agtcccgagct ggcctcagcc
 185461 accagtgagg actlaggcca ccoactaggag ctgcagccca gcgcaggcca cgcgcaaccct
 185521 aaggccacac agctcccagg cctgtlcccg agtccactca cactgacgat ggtccaaacc
 185581 acaatgagaa cgtctcgagg agtcaaataa gaaaggtggg aatatgttga aagaatggca
 185641 ttttcaaat tcttttgctt ggcctgcagg aaaagcgcaac aaaaatgagt aactctggcc
 185701 ttggaatcac agtttggctt aaatactgct tcgaaagctt gaagcagaact atctcgaact
 185761 tgcgagcttc agtttatatc cctgtagaag ggggatgatg tccggatagt ttccgtaggg
 185821 tgcgtgagc attaaacacg caactgggct acagggggct cagctgtgag cctggcacac
 185881 aatgttaata ggaacacagt caactgcaac ttaccatcat taacttttac gccaaaaggc
 185941 ttctcactag gtttttaagt ggcattgaaa aaccaagtga aattagcagc cctaacacgc
 186001 tgcccaggag taactggaaa actcttccac agcaacgtga ggcaacaacg cactctgtgc
 186061 agcctgttct gctcactgct tgtgcaagg gacattttgc actcaacctc agcagcctg
 186121 cctcgctctg gatattatta tgaataactt ccgtctagct actcaacctc tgaggacagc
 186181 agtctccac ctgtaaccag ggtgagatg cctcccagc caggcgctgt tgaggacagc
 186241 aggtgcccct tgtaaaagcc caggtcccta gcaaaagagc ctatcgtgtt ggcagcagag
 186301 gggattttaa cagtaggaca gacacatacg cagcaacaaa gtcagaaaagc gaagaaagag
 186361 cgcggggcgc gtgactgagg cctgtaatcc cagcacttta ggaagccaaag attgggcagat
 186421 cggctttagt tgggagtgtt agacacagct gatcaaaaag agcactttag caggcgaggc
 186481 gaaggaagga aggaaggaag gagaagagaa gaaaaagagg acaagcgaaa agaaaagagg
 186541 ctgctgcagc tggctctatc ctgtaatccc agcactttag gaggcgaggc aggcagcagat
 186601 acaaggtcag gagattgaga ccatctctgc taacacggtt aaacaccgtc tctactaaaa
 186661 atttttaaaa aataaaaaaa aaattagctg cactgtgttg cggcgccctg tgcctctgc
 186721 tactttggag gctgaggcag gagaattggc tgaacccgga aggcagagcgt cttcgactga
 186781 gctgagatcg gccactgca ctcagcctgt ggtlaataag tgaactccg tctcaaaaaa
 186841 aaaaaaaaaa aagaaagaaa acagaaaccaa tgggtgcttc ccaactgtcca aggccattat
 186901 tcagtgacag catccaatgc cgtcaggcca cgtcggctg cctcttccaa ccccaattatt
 186961 ctctccagct ctgtctggta catgggacac ttccagcagg cttgtccctgcc aggcacattt
 187021 cacacctggt acctctgcac acgtctctcc ctctgccggg gttgccaccg tgcgtttttt
 187081 ccagctggcc actcattctt ccagccccca ttcaaacagc cccccaggag cacactctcc
 187141 aagctctgccc cagagtgctc gtcagccacc ctgcagcagc gagccttggg tggctaccag
 187201 acctgcctca ttgctgacca tcacaggggga taacatcatg agaaggtgct gggctgaaaa
 187261 tgaatcacca atgaaactgc tgaatgaatg aaacagtgc aagtctctcc aaataagctc
 187321 tccatcagga atgcatagta cacaggtgtt gcccaccaa cagactgcga ccccatgtca
 187381 actccaccca ccttacctca ccttaagtcc gcagcaaaaca tctcttttca tttaggaat
 187441 taacctccag cgtaaaaatg gacacacctc tgaggccagc gatgtctcca gagacagtgt
 187501 gcgatgtcaa gggctcacca ttttaaccca gctctccag ggaattttac agttgtttgc
 187561 aatatggcca agttctaact aaatccagcc aaaggcagtg attttccaaa gcaacaatgt
 187621 atatttttct gtctccata gaaagaatga cagttgttgc agagccctgg tcaattggcc
 187681 gggctctcag actgtaggat gagcctcagc tgcccagta cgggcgtgct caggggccac
 187741 gggggctggg agagtggcca ctactcggg gcaggagttc cttagtagtc aattagtggc
 187801 gaccccatag tggggcgctt cagatgtctg gccagcacag ccagcctgac caccacacac
 187861 ccaatccaaa atagaaaact gcaagtggct ctgtctgtg ccaagcgccc gaggcccgca
 187921 cagagcaacc caaatgtcag cctcctccct cagcatcgat ccagcgtctc cctgaggccc
 187981 cctgtgctg agccacggtc cctcctggcc tctccaggaa gcagagcttc ctgacagctc
 188041 agtcacagga ctagagagcc gcaggcagtc tcaaaaggga tcccctagcc tctctctgtt
 188101 ccttcaggcg ctgcccccaa aagatcgagc agctcatggg cactagagct ttatttcca
 188161 aatgcacccc tcagctcaaa atagagaaaa ccaacttttc aaagaaccag ctggagtcca
 188221 acaggaaacc acagcagaaa aaccaatgat tcccgtcag ccaaggggca cgtggctaac
 188281 tggcatttct agcatcactg ccagctccctt gggcttgatg taccgcagc agaggaaaac
 188341 ggaagactga ctgcccagcc ataagtcacac cacacacata tccgcagat atgcacaagt
 188401 accgacacaa actcaacaga tgctcaaacac actgagtgc agaaaaccag acacatagc
 188461 gcacaagaga cactggcaca cgcacatgtt cgcacaggtg ccaggcgacc aattagcag
 188521 tacacaccaa cgttccacaa cacatgtgca cacactcaca aaggtacaca cacatacac
 188581 gcccaacccc cggatgcaag ccaagagactt gttcaggcag gttcaggcag caggcgccag
 188641 ctctcgctgt ctgagccctt tggagaccgg ggcgtccccc atgcataaatt cggccccctg
 188701 tcagaccagct caaaaacaccc tgccttcagc ccaagccctt ccaagttgagc cagtgctca
 188761 caaagcagcc ggaagggcgc ggaagtcctt cagcgcaggc cttgcccctc caggagactc
 188821 ctagggttaacc cttttccaaa acccagggga gagcagggga ctagagaagt ctatgaga
 188881 gcgtactcac cagcccatcg cagtgtctga gcgcacaga ggaagctctg gctagccgg
 188941 ccagctctcc gacgtagaga cagctccgga gcaggggctc cagcgggctg gtcgcttct

```

189001 cgcctcgcca ctccatagtg gcccccggcg ccacgaggcg ggcgttgggc cgcagccgca
189061 ggtgcaggtc tcggccaaag accgtgacat ttagaagagc gtgRtcgcca ggctcctcct
189121 cgttcgctcc gggRaagctc ggggtccgga ccggggcgcc cctgcctcct cgtacctcct
189181 ctctggaagc agctcgcgac accaYglggg acaccaagcg gcctcggggc tcagtgccga
189241 cgggcaacgc caggatgcgc tcgcclccgt gccccagggg ccgcctcgca ccgggaaggc
189301 gcgtagatc ggcggagacc acggagcccc agtgccctag agaccgccg gcaagccagc
189361 cccccccaga cctcccccga ctgcgaaggc aaggggcatt ccgccagcg accccagaag
189421 ceagcctgca cctccccgcg ttctctgcaa

```

[0254] Following are human cDNA sequences for transcript variant 1 (long form) and transcript variant 2 (short form) of *ADAMTS2* (cDNA sequences 1 and 2, respectively). Alternative splicing of the *ADAMTS2* gene generates two transcript variants, therefore, *ADAMTS2* exists in two forms: a "long" form comprising a molecule approximately 130 kDa in length, and a "short" form comprising a molecule approximately 70 kDa in length.

ADAMTS2 cDNA Sequence 1 (SEQ ID NO. 2)

NM_014244 Homo sapiens a disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motif, 2 (*ADAMTS2*), transcript variant 1, mRNA

```

1 atgattccgc cggcgggagc cgtctgccgc ctgctctgcc ccgcgctgct gctgctgctg
61 ctgctgctgc cgcgcgcgct cctgcgcgcy ccgcgcgcgc ccgcgaacgc caggctcgcg
121 cgcgcgcgcy acccccacag cgggcccctg gggcacggag cggagcgcat cctgcggctg
181 cccgtgcgca ctgacgccca gggcgcgttg gtgtccacag ttgtgtcggc agctacgctc
241 agagcagggg tacgagcccc caggggccgc ccgggtccga ccccgagctt cccgaggagg
301 aacgaggagg agcctggcag tcacctcttc cgtctcttgc ccgagacctg
361 cactgcgggc tgcgcgccaa gccccccttc gtggcggccc gggccactat ggagtggcag
421 ggcgagaagg gcaccaccgc cgtggagccc ctgctcggga cgtgtctcta gctcggagac
481 gtgcggcgcc tagccgaagc ctctctctgt gcgctcagca actgcgatgg gctggctggt
541 ctgtccgga tggaggagga ggaagtcttc atcgaaacct tggagaaggg gctggcgcg
601 caggaggctg agcaaggccc gtgcatgtgt gtgtatcgcc ggccaccacc gtcctcctct
661 ctcggggggc cacaggccct ggacacaggg gcctcctctg acagcctgga cagcctcagc
721 cgcgcctcgt gcgtctctaga ggagcacgac aacagctcga ggcggaggcg acgcaggcat
781 cgtgcagacg atgactacaa catcgaggtc ctgctggcggt tggatgactc ttgtgtgtag
841 ttccacggga aggagcacgt acagaagtac ctgctgacac tcattgaacat tgctaatgaa
901 atctaccatg acgagtcctt ggggtgccac atcaacgtgg tctgtgtcgc gatcatccct
961 ctgagctatg gaaagtccat gagcctcatc gagatcggga acccctctca gagcctggag
1021 aatgtctgcc gctgggccta cctccagcac aagccagaca cgggccagga tgaataccac
1081 gatcacgcca tcttctctcac acgcgaggac ttggggcctt ccggcatgca aggctatgct
1141 cctgtcacgc gcatgtgcca tcgggtccgc agctgcaccc tgaacctgat ggaagcgttc
1201 tctctcagct ttgtggtggc ccatgagact ggccacgtgc tgggcatgga gcacgacggg
1261 cagggcaacc gctgtggcga cgaggtcgcg ctggcgacga tcatgcccgc cctgggtcag
1321 gcgcctctcc accgcttcca ctggtcccgc tgcagccagc aggaagctgag ccgctacctg
1381 cactcctatg actgcctgct ggatgacccc ttgcgccagc actggcgccg cgtgccccag
1441 ctccccggag tgcaactact catgaacgag caatgcgcgt ttgacttcgg cctggggtcat
1501 atgtgtgca cggcgttcgc gacctttgac agctgcacag agctgtgggt cagcactcct
1561 gcaaacccct acttttgcaa gaccaagaag gggccccctt tggacgggac cactgtacatc
1621 cctggcaacc atgtgtttaa aggcacatgc atctggtgga cactgtacac cctcgaacac
1681 cagcgacgtg ggggcgcttg gaagtccgtt gcctcctgct cactgtacctg ttgcaacggc
1741 gtaagattca ggaccgcgca gtgtgacaac ggacaccggt ccaagcgccg ccgcatcctg
1801 tgcggccttg cctacgactt ccagctctgc agccgcaggc actgccccga cctcttggtt
1861 gacttcgcg aggagcagtg ccgcgagtg gacettgtact tcgagcacgg cgagcccgag
1921 caccactggc tgccccacga gcaccgggat gcccaaggaga gatgccacct gtactgcgag
1981 tccagggaga ccggggaggt ggtgtccatg aagcgcgatg gcatgatgg gacgcgtgag
2041 tctctaaagg acgcttccag cctctgtgtg cgcggggact gcaggaaggt gggctgtgac
2101 ggtgtgtatg gctccagcaa gcaggaagac tgtctggcgg ggaacacagc

```

```

2161 cactgcgaag tggctcaagg cactgtcaca cggtcaccca agaagcatgg ttacatcaag
2221 atgtttgaga tccctgcagg agccagacac ctgctcattc agggagtaga cgccaccagc
2341 caccatctgg ccgtcaagaa cctggagaca ggcaagtcca tcttaaatga agagaatgac
2401 gtggatgccca gtccaaaac ctctcatgcc atggcgctgg agtggagtaga cagagacagag
2461 gacggccggg agacgctgca gaccatgggc cccctccacg gcaccatcac cgttctggtc
2521 atcccgctgg gagacacccg ggtctcactg acgtacaaat acatgatcca tgaggactca
2581 ctgaatgttc atgacaacaa cgtccctggaa gaggactctg tggctatcga tggggcctcg
2641 aagaagtggt ctccgtgctc caagccctgt ggcggagggt cccagttcac caagtatggc
2701 ggctggacca gctgggacca caagaatgga caccgtggct tctgtgtccg cctctcgagg
2761 cccaaagcca tccgcagagc gtgcaaccca caggaatgct cccagccagt gtgggttaca
2821 ggcgaatggg agcactgtag ccagacactg ccagacaccc agggcgagag gcatggcagg gactccgtg
2881 cgctgcattc agcgcgacca acccgctccg tgcacgccaa gcatgtcaat tctgtccgga
2941 gacgcggcgc ggcgagagcg ccgggctctg agtcgcagc agcccgagga gcggccagtg
3001 cccgtggcga cgcgcagatg ctccagtaac tctgcccagg aggagcgatg tgagacagtg
3061 aggaactgca cgcgtggccc ctgtcccaga aacatctcag atccctccaa gaagagctac
3121 gtagttaagt ggcgtgcccg ccgggacccc gactcgccca tccggaaagt ctgcgaaag
3181 ggcactgccc aaggcgacaa gtcaatattc tgtaggatgg aagtctgttc cgcgtattgc
3241 tccatccagg gctacaaaca gctgtctcgc aagtctctga acctgtacaa caacctcaac
3301 aacgtgcgag caggataga gccaccgctg ggaagacaca acgacatga cgtgttcatg
3361 cctacccttc cagtcccac ttagcccatg gaggtgcggc catcaccaga acccccctg
3421 gagtctcttc tcaatgcctc cagacccaat atccacagag atcaacacaa caacattgac
3481 ctgatgatga cctacaaat ccatggcctg gaaatgaag tccagccacc caacctaatc
3541 cgtgacgac cgagccccta tgaaaagacc agaaacaaa gaatccaaga gctcattgat
3601 gagatgcgga agaaagagat gctcggaaag ttctaa

```

ADAMTS2 cDNA Sequence 2 (SEQ ID NO. 3)

NM_021599 Homo sapiens a disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motif, 2 (ADAMTS2), transcript variant 2, mRNA

```

1 atggatccgc cggcgggagc cgtctgcgcg ctgctctcgc cgcgctgctg gctgctgctg
61 ctgctgctgc cgccgcgcct cctgcgcgcg ccgcgcgcgc ccgcgaagcg caggctcgcg
121 gcgcgcgcgc acccccaggg cgggcgcctg gggcacggag cggagcgcat cctggcgctg
181 cccgtgcgca ctgacgccca gggccgcttg gtgtccacag tgggtgcggc agctaacgtc
241 agagcagggg tacgagcccc cagggcgcgc ccccgagctt ccccgagctg
301 aacgaggagg agcctggcag tcaactcttc tacaatgtca cgtctcttgg ccgagaactg
361 cactgcgcgc tgcgcgccc aagccgcctc gtgggcgcgc ggcgcctcat ggaagtgcag
421 ggcgagaagg gcaacagcg cgtggagccc ctgctcgagg ggcctcagca cgtcgatgat cgtcgagagc
481 gtggccggcg tagccgaagg ctagctcttc atcgaaacct tggagaaggg gctggcgctg
541 ctgactcgga agcaaggcgc tgtcatctgt gtatctgcgc ggcacccac gtcccctct
601 caggaggctg acgaaggcgc ggaacagggg gctcctcgg ggcagagggc agcagagctg
661 ctgcggggcg cacaggccct ggacacaggg cctcctcgg ggcgagggc agcagagctg
721 cgcgcgcctg cgtctctaga ggagcacgcc ctgctgggct cgtgtggcag tggatgactc tgggtgcag
781 gctgcgagc atgactacaa ctctgaggtc ctgctgacac tcatgaacct tcatgaatga
841 ttccacggga aggagcacgt acagaagtac ctagctgacg atcaacatgg tctgtgtgcy gatcactc
901 atctaccatg acagctctt gggtgccac agactctgag gagctcgga accctctcga gagctcgagg
961 ctgagctatg gaaagtcct ctagctctat agactcgga agcccgacga tgaataccac
1021 aatgtctgcc ctggggccta cctccagcag tttgggctct ccgcatgca aggctatgtc
1081 gatcacgcca tcttctctac accgcaggac atgtgcctgc ttgacaccca agctgcaccc
1141 cctgtccacc cagtgtgcca tccggtccgc ggcgcagctg tggcatgga cgcgcaggg
1201 tctctcagct ttgtgtgtgc ccatgagact ggccactgct tggcatgga cgcgcaggg
1261 cagggcaacc cgtgtgagca cgaagtgcgg ctggcgacca tcatgtgcgc cgtgtgcag
1321 gccgccttcc accgcttcca ctggtccgcg tgcagccagg aggaagctga cgcgtactgt
1381 cactcctatg actgctctg ggatgacccc tctgcacacc actggcgcg cgtgcccg
1441 ctcccgggac tgcactactc catgaacgag caatgcctct ttagcttcag cctgggctac
1501 atgattgtga cggcgttccg gacctttgac cctcgcaagc agctgtgtgt cagcactatg
1561 gacaaccctc acttttgcaa gaccaagaag gggcccccct tggacgggag tatgtgtgca
1621 cctggcaagt tcaaggccgg cgcggtggct catgctctgt atccagcac tttggaggag
1681 caaggtaggt ggaatgcctg a

```

[0255] Following are human polypeptide sequences for isoform 1 (long form) and isoform 2 (short form) of *ADAMTS2* (amino acid sequences 1 and 2, respectively).

ADAMTS2 Amino Acid Sequence 1 (SEQ ID NO. 4)

NP_055059 a disintegrin and metalloprotease with thrombospondin motifs-2 isoform 1; procollagen I N-proteinase; Procollagen N-endopeptidase [Homo sapiens]

MDPPAGAARRLLCPALLLLLLPPPLPPPPPPANARLAAAADPPGGPLGHGAERILAV
PVRTDAQGRVLVSHVSAATSRAQVRRARRAAPVRTPSFPGGNEEEPESHLYFNVTVFGR
DLHLRLRPNARLVAPGATMEWQGEKGTRVEPLLGSCLYVGDVAGLAEASSVALSNC
DGLAGLIRMEEEFFIEPLEKGLAAQAEQGRVHVYVRRPPTSPPLGGPQALDGTASLDS
LDSLRALGVLEEHANSRRRRARRHAADDDYNIEVLLGVDDSVVQFHGKEHVQKYLTL
LMNIVNEIYHDESLGAHINVVLVRIILSYGKSMSLIEGNPSQSLNVCRWAYLQKKPD
TGHDEYHDHAIFLTRQDFGPGSMQGYAPVTGMCHPVRSCITLNHEDGFSSAFVVAHETG
HVLGMEHDGQGNRCGDEVRLGSIAPLVQAFAHRHFWSRCSQQLSRYLHSDCLLD
DPFAHDWPAALPQLPLGLHYSMNEQCRDFGLGYMMCTAFRTDPCKQLWCSHPDNPFY
CKTKKGPPLDGTMCAFGKCHFKGHCWLTDPILKRDGSGWAWSPFGSCSRTCGTVGKF
RTRQCDNPHANGRTCSGLAYDFQLCSRQDCPSLADFREEQCRQWDLYFEHGDQ
HHWLPHHEHRDAKERCHLYCESRETGEVSMKRMVHDGTRCSYKDAFSLCVRGDCRK
VGCDGVIGSSKQEDKCGVCGGDNHCKVVKGTFTSPKKHGYIKMFEIPAGARHLIIQE
VDATSHHLAVKNLETGKFILNEENDVDASSKTFIAMGVWEYRDEDDGRETLQTMGPLH
GTITVLVIPVGDTRVSLTYKMIHEDSLNVDDNNVLEEDSVVYEWALKKWSPCKPCG
GGSQFTKYGCRRLDHKMVRHGFCAALSKPKAIRACNPQECSPVWVTGEWEPSCQ
TCGRGTGMQVRSVRCIQPLHDNTRSVHAKHCNDARPESRRACSRCLCPGRWRAGPWS
QCSVTGNGTQERPVP CRTADDSFGICQEEPRPETARTCLRGDPCPNISDPSKKSYYVQW
LSRPDPDSPIRKISSKGHCQGDKSIFCRMEVLSRYCSIPGYNKLSCKSCNLYNNLTNVEG
RIEPPPGKHNDIDVFMPITLPVPTVAMEVRPSPSTPLEVPLNASSTNATEDHPETNAVDEP
YKIHGLEHDVQPPNIPRRPSPYEKTRNQRIQELIDEMRKKEMLGKF

ADAMTS2 Amino Acid Sequence 2 (SEQ ID NO. 5)

NP_067610 a disintegrin and metalloprotease with thrombospondin motifs-2 isoform 2; procollagen I N-proteinase; Procollagen N-endopeptidase [Homo sapiens]

MDPPAGAARRLLCPALLLLLLPPPLPPPPPPANARLAAAADPPGGPLGHGAERILAV
PVRTDAQGRVLVSHVSAATSRAQVRRARRAAPVRTPSFPGGNEEEPESHLYFNVTVFGR
DLHLRLRPNARLVAPGATMEWQGEKGTRVEPLLGSCLYVGDVAGLAEASSVALSNC
DGLAGLIRMEEEFFIEPLEKGLAAQAEQGRVHVYVRRPPTSPPLGGPQALDGTASLDS
LDSLRALGVLEEHANSRRRRARRHAADDDYNIEVLLGVDDSVVQFHGKEHVQKYLTL
LMNIVNEIYHDESLGAHINVVLVRIILSYGKSMSLIEGNPSQSLNVCRWAYLQKKPD
TGHDEYHDHAIFLTRQDFGPGSMQGYAPVTGMCHPVRSCITLNHEDGFSSAFVVAHETG
HVLGMEHDGQGNRCGDEVRLGSIAPLVQAFAHRHFWSRCSQQLSRYLHSDCLLD

DPFAHDWPALPQLPGLHYSMNEQCRFDGFLGYMMCTAFRTFDPCKQLWCSHPDNPYF
CKTKKGPPLDGTMCAPGKFRPGAVAHACYPSTLGGQGRWIA

[0256] Modifications may be made to the foregoing without departing from the basic aspects of the invention. Although the invention has been described in substantial detail with reference to one or more specific embodiments, those of skill in the art will recognize that changes may be made to the embodiments specifically disclosed in this application, yet these modifications and improvements are within the scope and spirit of the invention, as set forth in the aspects which follow. All publications or patent documents cited in this specification are incorporated herein by reference as if each such publication or document was specifically and individually indicated to be incorporated herein by reference.

[0257] Citation of the above publications or documents is not intended as an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents. U.S. patents and other publications referenced herein are hereby incorporated by reference.

What is claimed is:

1. A method for identifying a subject at risk of osteoarthritis, which comprises detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the one or more polymorphic variations are detected in a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence in SEQ ID NO: 1-3;
- (b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;
- (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;
- (d) a fragment of a nucleotide sequence of (a), (b), or (c);

whereby the presence of the polymorphic variation is indicative of the subject being at risk of osteoarthritis.

2. The method of claim 1, which further comprises obtaining the nucleic acid sample from the subject.

3. The method of claim 1, wherein the one or more polymorphic variations are detected within a region spanning chromosome positions 178746000 to 178751000 in human genomic DNA.

4. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions selected from the group consisting of 210, 3608, 3609, 4318, 5593, 5629, 5639, 5640, 8943, 17968, 19887, 21034, 21085, , 21596, 23379, 23432, 24007, 26121, 26273, 26755, 27411, 27710, 27842, 28379, 29603, 31232, 31504, 32583, 32794, 32840, 33044, 33150, 33218, 33513, 33959, 34486, 36289, 36570, 38247, 38477, 38518, 38529, 38667, 39781, 39856, 39927, 40506, 41869, 42452, 44788, 46059, 46846, 47712, 48796, 49441, 49602, 49723, 50050, 50171, 50477, 50818, 50833, 50881, 50882, 51386, 51534, 52317, 52368, 52970, 53023, 53356, 53882, 54553, 55475, 55530, 55691, 55848, 55879, 56316, 56911, 57320, 57391, 57437, 57478, 57500, 59111, 59333, 59715, 59804, 59851, 59929, 60052, 60240, 60359, 60381, 60456, 60724, 60875, 60968, 60978, 60998, 61557, 62091, 62645, 62943, 63131, 63145, 63406, 63427, 63554, 63661, 64093, 64153, 64409, 64544, 65257, 65626, 65739, 66392, 66720, 69177, 69336, 69636, 69823, 69928, 70547, 70633, 71805, 72181, 72200, 72474, 72567, 72973, 73468, 73889, 75730, 75970, 76114, 76342, 76449, 76465, 76791, 78042, 80758, 80778, 81356, 81576, 81689, 81759, 81950, 82562, 83591, 83700, 83821, 83842, 83923, 83929, 84021, 84175, 84417, 84747, 85746, 86129, 86335, 87315, 87648, 87764, 87770, 88221, 90474, 91148, 91150, 91160, 91733, 91772, 91785,

93140, 93148, 96080, 96157, 96313, 96759, 97026, 97320, 97732, 98713, 99707, 99959, 100009, 100020, 100065, 100086, 101270, 101276, 101371, 101376, 101439, 101820, 102392, 102602, 102604, 102896, 189104, 189134 and 189205.

5. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in SEQ ID NO: 1 selected from the group consisting of 5640, 33150, 38247, 38529, 46846, 49723, 50050, 63427, 73889, 189104 and rs428901.

6. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in linkage disequilibrium with one or more positions in claim 3, 4 or 5.

7. The method of claim 1, wherein detecting the presence or absence of the one or more polymorphic variations comprises:

hybridizing an oligonucleotide to the nucleic acid sample, wherein the oligonucleotide is complementary to a nucleotide sequence in the nucleic acid and hybridizes to a region adjacent to the polymorphic variation;

extending the oligonucleotide in the presence of one or more nucleotides, yielding extension products; and

detecting the presence or absence of a polymorphic variation in the extension products.

8. The method of claim 1, wherein the subject is a human.

9. The method of claim 8, wherein the subject is a human female.

10. The method of claim 8, wherein the subject is a human male.

11. A method for identifying a polymorphic variation associated with osteoarthritis proximal to an incident polymorphic variation associated with osteoarthritis, which comprises:

identifying a polymorphic variation proximal to the incident polymorphic variation associated with osteoarthritis, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-3;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation;

determining the presence or absence of an association of the proximal polymorphic variant with osteoarthritis.

12. The method of claim 11, wherein the incident polymorphic variation is at one or more positions in claim 3, 4 or 5.

13. The method of claim 11, wherein the proximal polymorphic variation is within a region between about 5 kb 5' of the incident polymorphic variation and about 5 kb 3' of the incident polymorphic variation.

14. The method of claim 11, which further comprises determining whether the proximal polymorphic variation is in linkage disequilibrium with the incident polymorphic variation.

15. The method of claim 11, which further comprises identifying a second polymorphic variation proximal to the identified proximal polymorphic variation associated with osteoarthritis and determining if the second proximal polymorphic variation is associated with osteoarthritis.

16. The method of claim 15, wherein the second proximal polymorphic variant is within a region between about 5 kb 5' of the incident polymorphic variation and about 5 kb 3' of the proximal polymorphic variation associated with osteoarthritis.

17. An isolated nucleic acid comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-3;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

(e) a nucleotide sequence complementary to the nucleotide sequences of (a), (b), (c), or (d);

wherein the nucleotide sequence comprises a polymorphic variation associated with osteoarthritis selected from the group consisting of a cytosine at position 5640, a cytosine at position 33150, an adenine at position 38247, a thymine at position 38529, an adenine at position 46846, a cytosine at position 49723, a cytosine at position 50050, a cytosine at position 63427, a guanine at position 73889, a thymine at position 189104, and an adenine at position rs428901.

18. An oligonucleotide comprising a nucleotide sequence complementary to a portion of the nucleotide sequence of (a), (b), (c), or (d) in claim 17, wherein the 3' end of the oligonucleotide is adjacent to a polymorphic variation associated with osteoarthritis.

19. A microarray comprising an isolated nucleic acid of claim 17 linked to a solid support.

20. An isolated polypeptide encoded by the isolated nucleic acid sequence of claim 17.

21. A method for identifying a candidate therapeutic for treating osteoarthritis, which comprises:

(a) introducing a test molecule to a system which comprises a nucleic acid comprising a nucleotide sequence selected from the group consisting of:

(i) a nucleotide sequence in SEQ ID NO: 1-3;

(ii) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(iii) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(iv) a fragment of a nucleotide sequence of (a), (b), or (c); or

introducing a test molecule to a system which comprises a protein encoded by a nucleotide sequence of (i), (ii), (iii), or (iv); and

(b) determining the presence or absence of an interaction between the test molecule and the nucleic acid or protein,

whereby the presence of an interaction between the test molecule and the nucleic acid or protein identifies the test molecule as a candidate therapeutic for treating osteoarthritis.

22. The method of claim 21, wherein the system is an animal.

23. The method of claim 21, wherein the system is a cell.

24. The method of claim 21, wherein the nucleotide sequence comprises one or more polymorphic variations associated with osteoarthritis.

25. The method of claim 24, wherein the one or more polymorphic variations associated with osteoarthritis are at one or more positions in claim 3, 4 or 5.

26. A method for treating osteoarthritis in a subject, which comprises contacting one or more cells of a subject in need thereof with a nucleic acid, wherein the nucleic acid comprises a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence in SEQ ID NO: 1-3;
 - (b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;
 - (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;
 - (d) a fragment of a nucleotide sequence of (a), (b), or (c); and
 - (e) a nucleotide sequence complementary to the nucleotide sequences of (a), (b), (c), or (d);
- whereby contacting the one or more cells of the subject with the nucleic acid treats the osteoarthritis in the subject.

27. The method of claim 26, wherein the nucleic acid is RNA or PNA.

28. The method of claim 27, wherein the nucleic acid is duplex RNA.

29. A method for treating osteoarthritis in a subject, which comprises contacting one or more cells of a subject in need thereof with a protein, wherein the protein is encoded by a nucleotide sequence which comprises a polynucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence in SEQ ID NO: 1-3;
 - (b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;
 - (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;
 - (d) a fragment of a nucleotide sequence of (a), (b), or (c);
- whereby contacting the one or more cells of the subject with the protein treats the osteoarthritis in the subject.

30. The method of claim 29, wherein the treatment comprises administration of an effective amount of a composition comprising an active *ADAMTS2* polypeptide or fragment thereof, wherein the polypeptide fragment is selected from the group consisting of: 252-1211, 253-1211, 254-1211, 255-1211, 256-1211, 257-1211, 258-1211, 259-1211 or 260-1211 of SEQ ID NO: 4.

31. The method of claim 30, wherein the polypeptide or fragment has biological activity.

32. A method for treating osteoarthritis in a subject, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the one or more polymorphic variation are detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-3;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

administering an osteoarthritis treatment to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

33. The method of claim 30, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

34. The method of claim 30, wherein the treatment is selected from the group consisting of administering a corticosteroid, a corticosteroid, a nonsteroidal anti-inflammatory drug (NSAID), a cyclooxygenase-2 (COX-2) inhibitor, an antibody, a glucocorticoid, hyaluronic acid, chondroitin sulfate, glucosamine or acetaminophen; prescribing a heat/cold regimen or a joint protection regimen; performing joint surgery; prescribing a weight control regimen; and combinations of the foregoing.

35. A method for detecting or preventing osteoarthritis in a subject, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-3;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

administering an osteoarthritis prevention or detection procedure to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

36. The method of claim 35, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

37. The method of claim 35, wherein the osteoarthritis prevention is selected from the group consisting of administering a corticosteroid, a nonsteroidal anti-inflammatory drug (NSAID), a cyclooxygenase-2 (COX-2) inhibitor, an antibody, a glucocorticoid, hyaluronic acid, chondroitin sulfate, glucosamine or acetaminophen; prescribing a heat/cold regimen or a joint protection regimen; performing joint surgery; prescribing a weight control regimen; and combinations of the foregoing.

38. A method of targeting information for preventing or treating osteoarthritis to a subject in need thereof, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-3;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-3;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

directing information for preventing or treating osteoarthritis to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

39. The method of claim 38, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

40. A composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and an antibody that specifically binds to a protein, polypeptide or peptide encoded by a nucleotide sequence identical to or 90% or more identical to a nucleotide sequence in SEQ ID NO: 1-3.

41. A composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and a RNA, DNA, PNA or ribozyme molecule comprising a nucleotide sequence identical to or 90% or more identical to a portion of a nucleotide sequence in SEQ ID NO: 1-3.

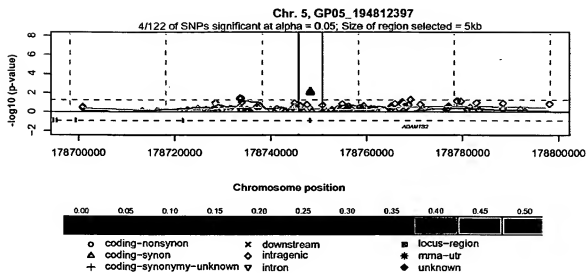
42. The composition of claim 41, wherein the RNA molecule is a short inhibitory RNA molecule.

Abstract of the Disclosure

Provided herein are methods for identifying a risk of osteoarthritis in a subject, reagents and kits for carrying out the methods, methods for identifying candidate therapeutics for treating osteoarthritis, and therapeutic and preventative methods applicable to osteoarthritis. These embodiments are based upon an analysis of polymorphic variations in nucleotide sequences within the human genome.

FIGURE 1

ADAMTS2 – DISCOVERY P-VALUES (female only)



Application Data Sheet

Application Information

Application Type::	Provisional
Subject Matter::	Utility
Suggested Group Art Unit::	Not Yet Assigned
CD-ROM or CD-R?::	None
Sequence submission?::	None
Computer Readable Form (CRF)?::	No
Title::	METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND TREATMENTS THEREOF
Attorney Docket Number::	524593008900
Request for Early Publication?::	No
Request for Non-Publication?::	No
Total Drawing Sheets?::	1
Small Entity?::	Yes
Petition included?::	No
Secrecy Order in Parent Appl.?::	No

Applicant Information

Applicant Authority Type::	Inventor
Primary Citizenship Country::	US
Status::	Full Capacity
Given Name::	Steven
Family Name::	MAH
City of Residence::	San Diego
State or Province of Residence::	CA
Country of Residence::	US
Street of mailing address::	12820 Via Nieve #74
City of mailing address::	San Diego
State or Province of mailing address::	CA

Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor
Primary Citizenship Country:: Germany
Status:: Full Capacity
Given Name:: Andreas
Family Name:: BRAUN
City of Residence:: San Diego
State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 3935 Lago Di Grata Circle
City of mailing address:: San Diego
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor
Primary Citizenship Country:: Germany
Status:: Full Capacity
Given Name:: Stefan
Middle Name:: M.
Family Name:: KAMMERER
City of Residence:: San Diego
State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 3825 Elijah Court, Unit 334
City of mailing address:: San Diego
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor
Primary Citizenship Country:: US
Status:: Full Capacity

Given Name:: Matthew
Middle Name:: Roberts
Family Name:: NELSON
City of Residence:: San Marcos
State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 1250 Calle Prospero
City of mailing address:: San Marcos
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92069

Applicant Authority Type:: Inventor
Primary Citizenship Country:: Sweden
Status:: Full Capacity
Given Name:: Rikard
Middle Name:: Henry
Family Name:: RENELAND
City of Residence:: San Diego \\\nState or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 7555 Charmant Drive, #1114
City of mailing address:: San Diego
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92122

Applicant Authority Type:: Inventor
Primary Citizenship Country:: United Kingdom
Status:: Full Capacity
Given Name:: Maria
Middle Name:: L.
Family Name:: LANGDOWN
City of Residence:: San Diego

State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 3701 Yosemite Street
City of mailing address:: San Diego
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92109

Correspondence Information

Correspondence Customer Number:: 25225

Representative Information

Representative Customer Number:: 25225